DOCUMENT RESUME

ED 181 746 FL 010 993

AUTHOR Dieterich, Thomas G.: Freeman, Cecilia
TITLE A Linguistic Guide to English Proficiency Testing in Schools. Language in Education: Theory and Practice, No. 23.
INSTITUTION ERIC Clearinghouse on Languages and Linguistics, Arlington, Va.
SPONS AGENCY Ford Foundation, New York, N.Y.: National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Dec 79
CONTRACT 400-77-0049
NOTE 61p.
AVAILABLE FROM Center for Applied Linguistics, 1611 N. Kent Street, Arlington, VA 22209 ($5.95)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *English (Second Language): Evaluation; *Item Analysis: *Language Proficiency; Language Skills; *Language Tests: Second Language Learning; Test Items; *Test Reviews: *Test Validity
IDENTIFIERS Metalinguistics: *Test Analysis

ABSTRACT

Part One of this guide explores issues in English proficiency testing. Tests are discussed in terms of the aspect of language tested, and of different kinds of test tasks. The following kinds of test task defects are treated: (1) tests that required literacy skills, (2) tasks that reduce to a vocabulary test, and (3) errant notions of linguistic complexity. Inherently faulty approaches to testing are discussed, including mimicry, testing passive comprehension, inferring lack of control from lack of performance, and using self-reported data. The value of specific test tasks (e.g., vocabulary test) in determining overall language proficiency is questioned. Linguistic artificiality in tests (e.g., demanding that answers be in complete sentences) is viewed as requiring metalinguistic knowledge from students. The implications of second language acquisition for language testing are discussed. Discrete point tests are contrasted with integrative tests, and the degree of correspondence between test content and order of acquisition of language skills is examined. Examples from specific tests are used throughout the guide. Part Two consists of an annotated catalogue of English proficiency tests. Each test is cross-referenced to relevant discussions in Part One. (JB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED181746

LANGUAGE IN EDUCATION: THEORY AND PRACTICE

23

A Linguistic Guide to English Proficiency Testing in Schools

Thomas G. Dieterich
Cecilia Freeman

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Center for Applied Linguistics
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

FL010 993

Published by
Center for Applied Linguistics

Prepared by
ERIC Clearinghouse on Languages and Linguistics

This work was supported by a grant from the Ford Foundation. We would like to thank Rosario Gingràs for organizing and nuturing the project in its early stages. Thanks also to Diane Bartosh (CAL publications program) for editing the final manuscript, and to Sonia Kundert for typing it.

This publication was also prepared with funding from the National Institute of Education, U.S. Department of Health, Education, and Welfare under contract no. 400-77-0049. The opinions expressed in this report do not necessarily reflect the positions or policies of NIE or HEW.

Language in Education: Theory and Practice Series ISBN: 87281-092-5

ISBN: 87281-110-7

December 1979
Copyright © 1979
By the Center for Applied Linguistics
1611 North Kent Street
Arlington, Virginia 22209

Printed in the U.S.A.

LANGUAGE IN EDUCATION: THEORY AND PRACTICE

ERIC (Educational Resources Information Center) is a nationwide network of information centers, each responsible for a given educational level or field of study. ERIC is supported by the National Institute of Education of the U.S. Department of Health, Education and Welfare. The basic objective of ERIC is to make current developments in educational research, instruction, and personnel preparation more readily accessible to educators and members of related professions.

ERIC/CLL. The ERIC Clearinghouse on Languages and Linguistics (ERIC/CLL), one of the specialized clearinghouses in the ERIC system, is operated by the Center for Applied Linguistics. ERIC/CLL is specifically responsible for the collection and dissemination of information in the general area of research and application in languages, linguistics, and language teaching and learning.

LANGUAGE IN EDUCATION: THEORY AND PRACTICE. In addition to processing information, ERIC/CLL is also involved in information synthesis and analysis. The Clearinghouse commissions recognized authorities in languages and linguistics to write analyses of the current issues in their areas of specialty. The resultant documents, intended for use by educators and researchers, are published under the title Language in Education: Theory and Practice.* The series includes practical guides for classroom teachers, extensive state-of-the-art papers, and selected bibliographies.

The material in this publication was prepared pursuant to a contract with the National Institute of Education, U.S. Department of Health, Education and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their judgment in professional and technical matters. Prior to publication, the manuscript was submitted to the Linguistic Society of America for critical review and determination of professional competence. This publication has met such standards. Points of view or opinions, however, do not necessarily represent the official view or opinions of either LSA or NIE. This publication is not printed at the expense of the Federal Government.

This publication may be purchased directly from the Center for Applied Linguistics. It also will be announced in the ERIC monthly abstract journal Resources in Education (RIE) and will be available from the ERIC Document Reproduction Service, Computer Microfilm International Corp., P.O. Box 190, Arlington, VA 22210. See RIE for ordering information and ED number.

For further information on the ERIC system, ERIC/CLL, and Center/Clearinghouse publications, write to ERIC Clearinghouse on Languages and Linguistics, Center for Applied Linguistics, 1611 North Kent Street, Arlington, VA 22209.

**From 1974 through 1977, all Clearinghouse publications appeared as the CAL-ERIC/CLL Series on Languages and Linguistics. Although more papers are being added to the original series, the majority of the ERIC/CLL information analysis products will be included in the Language in Education series.*

CONTENTS

INTRODUCTION by Bea Arias vii

PART ONE 1

ENGLISH PROFICIENCY TESTING IN SCHOOLS 1

Kinds of Language Tests 1
What Tests Test: Aspects of Language 3

Phonological Characteristics of English 3
The Vocabulary of English 4
The Syntactic-Semantic System of English 4
The Use of English in Different Contexts 4

How Tests Work: Kinds of Language Test Tasks 5

ISSUES IN TEST EVALUATION 5

DEFECTIVE TEST TASKS 7

Some Defects of Item and Task Design 7

Tests that Require Literacy Skills 7
Tasks that Reduce to a Vocabulary Test 8
Notions of Linguistic Complexity 8

Tasks that are Deficient in Principle 10

Mimicry 10
Passive Comprehension 11
Inferring Lack of Control from Lack of Performance 11
Self-Reported Data 11

DO SPECIFIC TEST TASKS DEMONSTRATE OVERALL LANGUAGE SKILL? 13

The Value of Vocabulary Tests 13
The Value of Phonology or Pronunciation Tests 15
Measures of Linguistic Creativity 16
Counting Linguistic Elements 18

LINGUISTICALLY ARTIFICIAL TEST TASKS 20

Understanding the Test Task 20
Ellipsis and Complete Sentences 21
Metalinguistic Instructions 23
Identical Test Items in Two Languages 25

IMPLICATIONS OF SECOND LANGUAGE ACQUISITION FOR TESTS 26

Types of Language Knowledge Tested: Discrete Point vs. Integrative Testing 27

Integrative Tests 28
Discrete Point Tests 28

Correspondence between Test Content and Acquisition Order 30

SUGGESTED READINGS 33

PART TWO 35

CATALOGUE OF TESTS 35

INDEX 53

INTRODUCTION

One of the educational developments witnessed in the 1970s, which has permeated every classroom and school district across the nation, has been the gradual increasing reliance on testing instruments rather than teacher judgments as sources of information for student placement and assessment. Indeed, teachers face a constant tension between validating their perceptions of their students' progress and that reported on standardized achievement tests. Until the testing developers can provide more curricularly tailored tests, teachers will have to deal with assessment procedures with which they may be unfamiliar and/or reluctant to implement.

This trend is most conspicuous in the areas of second language instruction, where diagnosis of student ability in various aspects of language proficiency has become an essential part of bilingual and English as a second language (ESL) programs. However, one of the most universal concerns for teachers in these programs has been their developing disaffection with assessment procedures and the student evaluations for which they are subsequently held accountable. Nevertheless, teachers have been called upon to administer measures with which they frequently are not familiar, which may not relate to the course of study as they have prepared it, and which may appear culturally or curricularly inappropriate.

The area of the assessment of oral language proficiency testing for limited and non-English-speaking students is one which has primarily developed in the last 10 years, paralleling the development of instructional methodology for bilingual education. It is noteworthy that when bilingual programs were first federally sponsored in 1968 there were very few tests which attempted to assess elementary students' command of spoken English. What was in existence primarily focused on assessment instruments for the acquisition of oral English in adults. In response to a demand which seemed to develop overnight, test developers struggled to provide instructional personnel with adequate measures. This created a situation where there were many newly developed tests on the market, but their appropriateness for assessing specific functions rarely emanated from teacher decisions, rather from decisions of psychometricians outside the realm of classrooms. Teachers then found themselves required to administer certain tests in which all of the following might obtain: (a) the children lacked test-wise skills, (b) the teachers were unfamiliar with the test, (c) the teachers had no input into test selection or administration, (d) the teachers could never use the test information collected, (e) the tests had little curricular relevance, and (f) the tests had little cultural relevance. In any of these situations, it is evident to see why disaffection on the part of instructional personnel with regard to assessment procedures and practices has developed.

The most efficient way to tackle this problem is for instructional personnel to develop competence in the areas of testing, testing development, and test interpretation. Teachers must know how tests are developed, what they can actually measure, what extrapolations from the tests are warranted, the deficiencies in tests, and, moreover, why it is necessary to use testing instruments for assessing students

rather than some other form of assessment procedure. Unfortunately, most teacher preservice programs cannot expand their already crowded curricula with additional courses on testing.

It is because the issue of assessment is so important and relevant to the issues confronted by today's educators that this monograph was developed. Essentially, it provides instructional personnel with the necessary background information for making informed choices in test selection and test interpretation. It addresses those issues which practitioners have consistently posed: *What do tests really measure? What type of student is this test appropriate for? Can I trust the test? Will my students be adversely affected by taking the test? What can I learn from the test about my students?* By providing the answers to these questions it is hoped that some of the mystique surrounding testing will be eliminated for instructional personnel. This monograph, then, is very timely in providing this basic introduction in test development and application.

This Guide also serves two other purposes: (a) it elaborates on the definitions for such concepts in bilingual and ESL programs as "language dominance" and "language proficiency" and (b) it reviews tests in an annotated format. Terms such as proficiency and dominance have been so popular in current programmatic use that they have escaped consistent definition. Language proficiency in one district may be language dominance in another. By providing an operational definition of language proficiency and contrasting it with the definition of language dominance, this book bridges the gulf between current research in second language acquisition and the application of this knowledge to the classroom.

Finally, this monograph reviews those tests which have met with the most widespread usage. In annotated format, each test is reviewed with a cross-reference to those specific areas where the test may be limited. This review is unique from others recently published in that it describes the tasks involved in the assessment procedure, or the stimulus to which the student is required to respond. This feature makes it very usable for instructional personnel by providing them with a better idea of what tasks are required for tests they may be selecting for programmatic use.

In conclusion, A Linguistic Guide to English Proficiency Testing in Schools fills a critical void in the professional development of teachers of limited and non-English-speaking students: It provides the essential background information on testing needed by today's educators, it addresses the latest issues in bilingual and ESL education, and it provides a very useful annotated list of tests in use. For the inservice and professional development of teachers, the book serves a practical and commendable function.

Bea Arias
September 1979

Part One

ENGLISH PROFICIENCY TESTING IN SCHOOLS

The student population in many U.S. school systems these days includes students from diverse language and cultural backgrounds, many of whom are native speakers of languages other than English, and who vary widely in their ability to use English in a school setting. Teachers and administrators in this situation have a recognized responsibility to identify students who have limited proficiency in English and find out how well each student is able to use English for placement in an appropriate instructional program. But, despite widespread agreement on these goals, no consensus currently exists on answers to certain crucial questions. *What sort of language ability is sufficient to enable a non-native speaker of English to get along in an English-speaking classroom, and what indicates a lack of that ability? How can we find out just how well a student can use English, at a given time?*

In spite of the present lack of consensus, there is some research which can form a preliminary basis for answers to these questions. Furthermore, there are many teachers and other people working in schools whose experience has provided them with at least partial answers. There is also a large and growing number of language assessment instruments being made available to schools, which promise to assist the schools in identifying and placing students of limited English-speaking ability. School systems, under pressure to demonstrate that student placement is based on some reliable and valid measure of language ability, frequently face the problem of choosing a test to use. This problem is compounded by the fact that there is a great deal of variation among currently available tests, in terms of language content, the types of language tasks involved in their use, scoring procedures, and ways to interpret scores.

This Guide reviews a number of these tests and reports on a linguistic evaluation of their design and content as well as introduces essential concepts in language testing. Since several publications are now available which provide psychometric data on tests, evaluation of their cultural appropriateness, and/or evaluation of their practical usefulness in a school setting, we will not include those considerations in our reviews. Instead, we will try to provide information about available tests which previous catalogues do not contain.

In one way or another, the tests we reviewed claim to assess students' ability to produce, understand, and/or use language correctly or appropriately. Linguistics, as a discipline, deals with the structure, use, and acquisition of language, and provides an important framework for evaluating whether, and how well, language tests do what they claim. The present review is based on an analysis of the language of test items, the nature of tasks in tests, and the manner in which responses to test items are to be judged or scored. Given this analysis, we evaluate tests in light of evidence from linguistics and related disciplines about how language works, how it is used and how it is learned.

Kinds of Language Tests

The language tests under discussion are generally used in schools to serve either or both purposes of: (a) identifying students with limited English proficiency and/or (b) placing students in an appropriate type or level of instructional program (e.g.,

bilingual instruction, ESOL, or a regular English-speaking classroom). Test developers make different claims for what type of information their tests will yield about students' language abilities. Most of the language tests we have considered are claimed to assess students' language proficiency, bilingual dominance, or both.

An *English language proficiency test* is designed to provide an answer to the question (roughly put): How well does the student know English? English proficiency--what it is to "know" English--is given different operational definitions in each theoretical, historical, and legislative context. Reflecting various notions of what constitutes language proficiency, different tests are designed to assess students' skills in certain aspects of language use. Most tests of language proficiency involve the comprehension and production of oral language. Some also include a component for testing written language skills. In any given school situation, the apparent value of a test will depend in part on whether the test effectively assesses that array of language abilities that the school deems crucial to students' success in an English-speaking classroom. Lately, educators have spoken of the "four skills" composing English proficiency: speaking, listening, reading, and writing. While we recognize that proficiency in each of these skills is ultimately essential for success in English-speaking classrooms, we intend to focus on the oral/aural skills of speaking and listening. These are primary, in that they constitute the basis on which the literacy skills of reading and writing must be built, and in that they must form the basis of any reasonable definition of language proficiency. An illiterate native English speaker is simply illiterate, not unproficient at English. Henceforward, when we speak of language proficiency, language dominance, etc., we will mean *oral* proficiency or dominance. The reviews that follow will, for the most part, cover only those tests or parts of tests which claim to assess speaking and listening ability. There, we will touch on some of the issues involved in testing particular language abilities or skills as a measure of oral language proficiency.

The question of *language proficiency* is frequently approached through the notion of *language dominance*. A student's "dominant" language is sometimes determined on the basis of use patterns (e.g., the language of the home, the language most often spoken, etc.) in accord with state and federal legislation, and this method (which often yields conflicting results) of determination is increasingly supplemented with a test of dominance. A test of bilingual dominance is generally designed to determine which of two languages students know better/less. Superficially, it might appear that dominance can be ascertained without simultaneously determining how well students *know* each language. For example, the dominant language could be determined without determining overall language proficiency by comparing students' control over some small sample of each language like a list of words. The trouble with this idea, however, is the inescapable possibility that students might perform a limited task (e.g., naming pictures) better in language A than in language B, but, under ordinary circumstances, may be better at speaking and understanding language B than language A. Unless the task in a dominance test yields an indicator of students' *overall ability* in each language, the test may fail to indicate language dominance (if, in fact, students even have a single dominant language). Thus, for language testing, the question of dominance is actually a question of students' relative overall proficiency in one language as compared to others. Some issues pertaining to dominance testing will be discussed later.

There are a few tests which claim to assess students' "fluency" or "language facility." However, *fluency* and *language facility* are given no technical definitions which differentiate them from language proficiency. The methods used to assess fluency or language facility are the same as those used in some language proficiency tests and, in practice, tests which claim to measure fluency or language facility appear to be used as language proficiency tests.

Some tests are presented as tests of "achievement." Of course, *achievement* can only be measured in terms of progress toward some specified goal. Generally, achievement

is measured in terms of mastery of specified curriculum objectives, as in an ESOL program. In these cases, the value of an achievement test as a measure of students' overall English ability depends partly on the value of the curriculum. A test based on a particular set of curriculum objectives will measure overall English proficiency only insofar as those objectives actually comprise what it is to know English, and those objectives reflect the order and patterns in which English is naturally acquired as a second language. A considerable amount of research indicates that children acquire the grammatical structures of English in a systematic, regular way, but the order of acquisition so far has been documented for only a limited number of grammatical morphemes. Research has not yet established an overall picture of the English acquisition hierarchy in sufficient detail to serve as the basis for an entire ESOL curriculum. On the other hand, due to the time lag which often separates research findings and their practical application, most ESOL curricula presently do not even reflect the considerable evidence which already exists for sequence in the acquisition process. In general, we can expect that achievement tests based on a list of specific curriculum objectives will be of limited value as tests of overall English proficiency.

Some tests, finally, claim to be "diagnostic." A test is *diagnostic* if it gives qualitative *and* quantitative information regarding students' language proficiency. A diagnostic test must do more than assign students to some category (e.g., one of several different levels of English proficiency) on the basis of how much English they know; it must also yield information regarding aspects of the English linguistic system students do or do not control. A diagnostic test will be more useful as a *guide* for language instruction because it determines precisely what students do not know, and only from that information can we decide precisely what needs to be taught. A test of language proficiency or bilingual dominance may or may not be diagnostic.

What Tests Test: Aspects of Language

A test uses a number of different methods to assess English proficiency, and each test's methods vary according to the component of language the test selects as representative of the linguistic system.

Every spoken language has a universal framework of properties, or components: pronunciation, grammar, vocabulary, forms of discourse, and rules for use. Together, these components comprise the linguistic system of a language, and the different aspects of the system form a hierarchy of levels so that units at each level are organized into larger units at the next level. Thus, the sounds of English are organized into words, words combine into grammatical structures which are organized to express various meanings, and utterances are used systematically in social situations.

In general, when language is learned in natural social situations (i.e., in the absence of formal instruction), the whole of a language system is acquired in meaningful "chunks" as communicative situations require or permit. In acquiring overall control of a linguistic system, the learner is gaining control of various subsystems which can be analytically separated: pronunciation (phonology); the grammar proper (syntax); the vocabulary (lexicon); patterns of discourse beyond the sentence; meanings associated with the grammar, vocabulary, and patterns of discourse (semantics); and the rules of use (pragmatics).

Phonological Characteristics of English. Differences between sounds may signal differences in meaning in some languages. Where sound distinctions function to differentiate meaning, speakers of the language tend to perceive "different sounds." For example, the sounds of *v* and *b* constitute a functional contrast in English because they serve to differentiate between the spoken words *drivel* and *dribble*, *vest* and *best*, etc. These two sounds do not distinguish between any such word pairs

in Spanish, and so are not functionally different in that language. Similarly, a native speaker of Japanese may have difficulty in hearing or producing the distinction between the sounds of *r* and *l*--which differentiates English words such as *right* and *light*--because this distinction does not serve to differentiate any words in Japanese. Conversely, there are sounds that do not differentiate any words in English, but do in other languages. The sound of *k* in *skull* is acoustically and perceptually different from the sound of *k* in *ski*, but this is a difference that is ignored in English. On the other hand, a Turkish speaker would quickly perceive the difference between the two sounds, because in Turkish these two sounds alone can serve to distinguish between different words. A person acquiring the phonological system of a new language must learn to ignore certain differences among spoken sounds, and at the same time to perceive and attend to certain other differences, which are important to understanding the language.

The Vocabulary of English. The language learner must learn the meanings of words, and how meaning is affected as words occur in different contexts. Words such as nouns or verbs might be learned explicitly as vocabulary items, but "function" words (such as articles, relative pronouns, and conjunctions), which indicate relationships among other words in a sentence, must be acquired as part of the process of learning to construct sentences.

The Syntactic-Semantic System of English. Learning the syntax and semantics of English involves learning how words are organized into meaningful sentences and discourse. The syntactic-semantic system distinguishes a mere list of words from a *grammatical, meaningful* sentence of English. In learning how to make an English sentence, the language learner must master the grammatical relations which hold within sentences. Students must learn, for example, (a) the different forms which pronouns take as subjects or objects (*he* vs. *him*, *I* vs. *me*); (b) the grammatical processes underlying the structure of sentences (forming questions, imperative or passive sentences); (c) how to form embedded constructions (e.g., relative clauses: *The man* [*who is*] *near the door is my brother* or sentential direct objects: *He expects me to read this whole book*, *He knows* [*that*] *we'll be late*); and (d) how to control grammatical relations which occur among different sentences in discourse (e.g., pronominalization or cases of ellipsis where *ten* and *I'm ten* [but not *am ten*] are grammatically acceptable responses to a question like *How old are you?*).

The Use of English in Different Contexts. English sentences must be more than meaningful and grammatical. They must be appropriate to the particular linguistic and social context in which they are used and communicate the speaker's intended meaning in that context. In learning the conventions which govern spoken English discourse, students come to understand how linguistic forms which are identical in content and structure can convey various meanings in different contexts.

Recognizing the interaction of language forms and context is a linguistic skill which is fundamental to understanding conversation and to being understood. Another related aspect of language use which must also be learned is how language can be used appropriately and effectively to perform different functions which are often highly specific to a given situation.

The interrelated aspects of language are all essential to the use of English, but there are important differences among them in terms of how they are learned and how they operate as part of the overall knowledge of English. There is research evidence that the acquisition of grammatical structure by children learning English as a second language is more systematic than the learning of other aspects of the language; there appears to be a regular pattern to the acquisition of English structures by children of diverse backgrounds and diverse English-learning experiences. In contrast, how children acquire phonology, vocabulary, and the functional use of English is more likely to depend on the learner's native language, culture, individual experience, and the nature of explicit instruction received.

The acquisition of English grammar, then, is the most regular, systematic, and predictable aspect of the language-learning process. This mirrors the fact that, among the various aspects of the English language system, the grammatical system is most uniformly shared by all English speakers. That is, the basic syntactic structures of the language are least likely to vary among speakers of different social class, age, sex, or ethnic or regional background, and are less likely to vary over time, or in different situations of use. Each of these factors, in contrast, tends to influence a speaker's pronunciation of the language, and the vocabulary controlled.

The problem for tests of language proficiency is how to gain access to students' knowledge of the overall linguistic system. In general, the strategy that test developers take is to evaluate students' control over one or two of the four aspects of the linguistic system, on the assumption that students' proficiency in those aspects of the total language system is representative of overall ability.

How Tests Work: Kinds of Language Test Tasks

Another point of difference among tests is their method of assessing students' control over the aspect of language which they select to measure. Different tests involve the student in one or more of the following types of tasks:

- Answering questions about pictures, about a discourse, or general questions.
- Describing, or telling a story about, pictures, objects, places, or people.
- Paraphrasing something which is said.
- Grammatically manipulating sentences--changing tense or number, conjugating verbs, changing sentence form, etc.
- Completing cloze passages or sentences.
- Repeating words, sentences, or stories.
- Recalling words from lists of words, generally presented in two languages.
- Discriminating between words.
- Pointing to or marking pictures, words, sentences, or objects which correspond in some specified way to an oral cue.
- Naming objects in pictures or in the physical environment.
- Performing commands.

All of these tasks involve students in oral language processing or production, or both. A few tests include tasks which also involve students in reading, as for example, a test which requires students to select, from among several written sentences, one which corresponds to an orally given sentence cue. Following, we will discuss and evaluate specific tests in terms of the aspects of language they seek to measure and how they seek to measure them, and in terms of the tasks they impose.

ISSUES IN TEST EVALUATION

The tests to be discussed claim to measure English language proficiency. The crucial question to ask of any English language proficiency test is whether it does what it claims to do, which is a question about test validity. In evaluating tests, the question of validity is by far the most important and complicated.

There are various ways to pose the question of test validity, and various ways to evaluate it. Most commonly, claims of test validity are supported by certain types of experimental evidence from studies conducted over a period of time. Studies conducted to validate language proficiency tests are commonly designed to demonstrate that (a) a test accurately predicts some independent, but presumably relevant aspect of students' future performance (e.g., subsequent scores on standardized achievement

tests or success in an English-speaking classroom as judged by a teacher) and (b) the test yields results which correlate highly with the results of some other tests or assessment procedures whose validity has presumably been independently established. Test makers may be expected to provide data from such studies in support of claims that a test validly measures language proficiency. When such data is provided, it should be carefully examined and evaluated before using the test.

Note, however, that this type of experimental evidence alone does not directly assess whether a test actually measures proficiency in or control of a given language. Rather, these prediction and correlation studies merely allow us to infer, with greater or lesser confidence, that the test actually measures language proficiency. Whether or not we can put faith in such an inference, in any one case, depends on the relevance of the validation studies to the assessment of language proficiency, on our confidence in the research design and procedures of the studies, and on the accumulation of empirical evidence from the studies.

Our review evaluates tests analytically rather than experimentally. We ask certain relevant questions of a test, and we seek answers by examining its design, content, and procedures, rather than by drawing inferences about language ability from experimental observation of behavior that (presumably) depends on that ability. What makes this approach relevant and essential is the fact that every English language proficiency test reflects--implicitly or explicitly--a particular notion of what it is to "know" English, and that notion in each case must be subjected to evaluation. A language proficiency test may be viewed as an instrument which sets up an operational definition for something called *language proficiency*. Since what counts as language proficiency might be conceptualized differently by different test makers, it is important to ask whether the notion of language proficiency embodied in a particular test is appropriate for a given purpose (e.g., for assessing how well a child speaks English in order to provide the needed instruction). Further, given a particular test which embodies a particular notion of what abilities constitute proficiency in English, it is important to ask how adequately the items on the test sample those abilities--how well performance on the test represents control of those abilities.

Answers to these questions emerge from linguistic (and psychometric) analysis of tests. In the case of language proficiency tests, linguistic analysis is a doubly appropriate tool for evaluation, for both the theoretical construct which the tests are intended to measure and the methods for measuring that construct are linguistic in nature. Our discussion is based on analysis of the language content of tests and of the nature of linguistic tasks involved in them. Given this analysis, tests are evaluated according to what is known about the system and structure of language, and about how children learn language. Thus, this evaluation rests on a theory-based understanding of how language works and of the language component the tests are designed to measure.

For purposes of this review, two assumptions are made regarding the question of test validity: (a) each particular task imposed is an appropriate one for assessing control of some specific aspect of language and (b) the demonstrated knowledge of those specific aspects of language covered by the test indicates the level of overall language ability. If either of the two assumptions fails to be true (wholly or in part), then this damages the test's claim to be a valid measure of overall language proficiency.

Our assessment of test validity consisted of a linguistic analysis of a test's item content (including the strategies for scoring performance) and a comparison of the overall design of the test with the major components of language as a structured system of communication. A linguistic analysis can, where appropriate, serve to raise suspicions about test validity by exposing cases of *confounding** (assumption

**Confounding simply means that a test item or task that has been designed to measure knowledge of, or skill in, some specific aspect of language, might, in reality, be*

"a") and by providing evidence that certain aspects of language frequently measured on proficiency tests do not provide good indices of *overall* capability (assumption "b"). For each validity criterion, we present examples drawn from tests now on the market or currently used in schools, of how particular test tasks can either succeed or fail.

Another area we examine concerning test validity (in a sense) is the extent to which a test claims to elicit "natural" language--language which meets the requirement of daily use. Performance on a proficiency test should be a measure of language competence and not merely a measure of test-taking ability. Tasks posed on language proficiency tests should--as nearly as possible--match the conventions and requirements of ordinary language use. Artificial rules and requirements that may serve the purposes of the test may also serve to make the test language unlike the language of everyday communication.

Finally, we raise a further issue in the evaluation of tests, one which has less to do with purely linguistic analysis than the first two, but which depends more on research in language acquisition. Assuming we know what features constitute important parts or aspects of language, how are these *acquired* by a second language learner? If there is a generalizable sequence of progress in learning English as a second language, how do we test where an individual stands in this sequence?

DEFECTIVE TEST TASKS

Some Defects of Item and Task Design

Tests that Require Literacy Skills. Several of the oral language proficiency tests we reviewed contain subtests that are explicitly designed to test reading and writing abilities. For example, one subtest of the Short Test of Linguistic Skills* is designed to assess English listening ability, but the tasks entail (a) selecting a word given orally from a written, multiple-choice list of three words; (b) writing a dictated word, phrase, or sentence; and (c) writing certain items which require English spelling and writing ability (e.g., *print the word animal and underline all the vowels*). The *Robstown Oral Language Inventory*--although a different kind of test--also has the same problem of confounding oral proficiency with literacy. This test requires students to answer questions about when and where they use English or Spanish (e.g., *What language do you speak at home? When you dream, what language is spoken in your dream?*). These questions are asked orally of younger children, but older students (Level 2 of the test) must read them, so obviously something more is required than the ability to understand and reply to spoken English questions.

Both tests also contain subtests of Spanish proficiency with the same confounding--Spanish reading and/or writing ability is required to perform the test task successfully.

By making these criticisms, we do not intend to diminish the importance of literacy skills in language proficiency; they are merely examples where literacy is included as a part of an assessment of something else. Because of the way the tasks are designed, the items are tapping more than *one* performance objective, and response to the items depends on the ability to perform in at least *two* skill areas.

---

measuring some different *knowledge or skill. Performance on a confounded item or test task might depend on one's skill in what the item or task was designed to test; however, it might also depend--wholly or in part--on one's skill in some different aspect of language.*

**All tests mentioned in this discussion are further identified in Part Two, where comments about specific tests may be found under the names of the tests, which are arranged in alphabetical order.*

Tasks that Reduce to a Vocabulary Test. Many tests include a task where students must identify a picture corresponding to a phrase or sentence which has been given orally. With proper design, this sort of format might be employed to test whether students can correctly process and understand the structure and meaning of a spoken expression. In many cases, however, students can choose the correct picture by understanding a single crucial word in the stimulus expression. In one item of the Language Assessment Battery, the oral stimulus is *The cat and three kittens are playing*. Of the three pictures from which to choose, only one contains a cat; the other two show different objects. The item does not test understanding of the stimulus sentence as a whole, because by understanding only the word *cat*, the correct picture could be chosen. In another item from the same test, the stimulus sentence is *The puppy is following the girl*. The three picture choices show a girl with a puppy in her arms, a girl with a puppy following her, and a girl with a puppy between her feet. By understanding the single word *follow*, the picture could be correctly identified. Note what might happen had one of the pictures shown a girl following a puppy: students would have to distinguish between a girl following a puppy and a puppy following a girl, and consequently would have to understand the stimulus *sentence*, rather than just one word.

The Shutt Primary Language Indicator Test raises very similar issues. On this test students identify pictures corresponding to what the examiner says. Some of the items contain extended expressions or complicated structures, like *What you would get if you asked for a vegetable*. However, students could possibly get this item correct by recognizing the single word *vegetable* and identifying the corresponding picture. This item, like a number of others on the Shutt test, may not necessarily assess comprehension of extended expressions, but may just assess vocabulary knowledge.

To identify test items of this sort, simply ask, "What kind of knowledge would be required to get this item right? Do students have to understand the entire expression, or is it sufficient to understand one or two words?" If it is enough to understand a word or two, then the item cannot claim to solely assess students' ability to understand an extended phrase or sentence.

The Crane Oral Dominance Test is another instrument that primarily measures vocabulary knowledge. The format of this test is unique among those reviewed, since it involves a memory task. Eight words are presented orally--four in each language--then students recall the eight words. If more Spanish words are recalled, students are considered Spanish dominant, and likewise for English. The test claims to assess what language students "think" in, but it seems clear that this is predominantly a vocabulary test, because students will have a much better chance of recalling known words than unknown ones.

It is relevant to ask whether or not it is bad for a test to measure only vocabulary. We are aware that there is much more to language than knowing words. The ability to understand and properly use extended expressions is--in some way--more central to language proficiency; however, the question of how well vocabulary tests measure overall ability remains. We will discuss this at greater length later, since this seems to be a controversial issue.

Notions of Linguistic Complexity. Many language proficiency tests require oral responses. Frequently these spoken responses are scored according to their "complexity." Students get more points--and a higher proficiency score--for producing "more complex" answers. Categories of linguistic complexity are provided as an aid to scoring. A typical hierarchy of categories is that of the Test of Grammatically Correct Spanish/English: clause fragment (one point); simple sentence (two points); simple sentence plus--includes compound subjects and predicates (four points); compound sentence (six points*); complex sentence (eight points*); and compound-complex sentence (ten points*).

**Personal communication from Margarita Lopez de Mestas, Coordinator, Bilingual Education Project, Las Cruces, New Mexico Public Schools*.

As an overall hierarchy of complexity, this seems reasonable from a linguistic point of view. In particular, it makes sense to say that the number of clauses and how they are arranged is one important factor contributing to the linguistic complexity of a sentence. However, there are also some problems with this hierarchy. Chief among them is scoring any clause fragment below any complete sentence. Some clause fragments are perfectly natural responses to questions, and require as much linguistic knowledge as a full sentence response. For example, *Playing cards with my friends* is just as natural a reply as *I am playing cards with my friends* in response to the question *What are you doing tonight?* Furthermore, one must know exactly where it is permissible to omit *"I am"* and where it is not. If the question had been *What's going on tonight?*, *"I am"* could not have been omitted. So, although the fragment response is superficially less complex than the complete sentence response, the appropriate use of fragments may indicate at least as much English proficiency as production of full sentences.

For the same reason, we should not score compound subject and predicate sentences as less complex than compound sentences: *John cooked the eggplant and Harry ate the eggplant* is no more complex linguistically than *John cooked, and Harry ate, the eggplant.*

Another problem with the above scoring procedure is that it is not always easy to separate simple from complex sentences, because it may not be a clear-cut decision whether or not a given sentence contains a subordinate clause. For example, some linguists regard all infinitive expressions as subordinate clauses and some do not; controversy surrounds this issue. Sentences that may contain subordinate clauses, but do not obviously do so, include: *John had his pocket picked last night, You should come see the new car I bought, A man happy in his work is a productive employee,* and *I anticipated John's asking for a raise.*

The Del Rio Language Screening Test uses another sort of hierarchy to determine sentence complexity: (a) sentences with transitive verbs are more complex than those with intransitives; (b) sentences with more auxiliary verbs are more complex; (c) questions are more complex than assertions; (d) negated sentences are more complex than positive ones. From a linguistic point of view, these factors may indeed contribute to complexity, but other important factors are ignored. For example, the failure to consider clause structure as contributing to complexity leads this test to claim that the following two sentences are approximately equal in complexity (though unequal in length): *I like ice cream, I want to go to school to see my friends.* On intuitive as well as linguistic grounds, the latter sentence is much more complex than the former. Tests that score students on complexity of response seldom consider factors beyond the length or structure of sentences.

The Ott Test of Oral Language provides a complexity hierarchy based on sentence structure. The scoring instructions state explicitly that *two* responses in a given category do not count the same as one response in a higher category. So, a response consisting of two prepositional phrases counts the same as a one-phrase response; likewise, two simple sentences score the same as one simple sentence, and therefore less than a complex sentence. In some cases, this scoring strategy will cause the wrong evaluation of students' language ability, because there may be inferential relations between two independent simple sentences that are as real and complex as the structural relations between two clauses of a complex sentence. For example, the statement *John went to the store because he wanted to buy some bread* expresses a causal relationship between two clauses. But the two sentences, *John went to the store. He wanted to buy some bread*, imply the same causal relationship, and merely represent the use of one linguistic mechanism rather than another to convey the information. There is no principled reason to regard the second example as less complex than the first.

Tasks that Are Deficient in Principle

The previous examples are cases of faulty items or task design which caused particular parts of tests to fail to reflect the structure or organization of language. There are, however, some test tasks which--by their nature--are of limited usefulness in assessing overall language proficiency.

Mimicry. Several of the tests we reviewed use an oral mimicry task. Basically, the task requires students to repeat exactly words, phrases, or sentences spoken by a test administrator. The obvious problem with mimicry is that it does not correspond to any skill necessary for daily use of language for communication. In the tests we reviewed, the mimicry task was used either as a means for eliciting a speech sample on which to assess pronunciation* or as a means of assessing ability to process, remember, or comprehend sentences.

The mimicry task cannot provide a direct measure of comprehension, since we can imitate--more or less precisely--any number of things that we cannot understand. The ability to mimic speech well is not the same as the ability to process and comprehend language; if it were, parrots could be proficient English speakers.

Tests that rely on pure mimicry to assess sentence-processing ability are relying on a connection between understanding and memory. In psychological experiments, it has been shown that a person can perform better at remembering a sequence of elements when those elements are organized into a single, larger unit. It is easier to remember a five-letter word than a meaningless sequence of five letters, and easier to remember a seven-word sentence than to remember those same seven words in scrambled order. According to this principle, students should perform better at remembering and repeating English sentences if they are understood as sentences rather than as meaningless groups of words or--worse yet--as meaningless sequences of sounds. Two sections of the Del Rio Language Screening Test apparently rely on this principle. This test states that "...a child usually cannot repeat a syntactically complex sentence which is beyond his linguistic competence..."; this is probably too strong a claim, but illustrates the test makers' adherence to the above principle.

There are three problems with this sort of test: (a) It is possible to mimic exactly sentences not understood or not understood well, and this factor cannot be ruled out in any particular student's performance. Furthermore, one student may be a better mimic than another and so will do better on this sort of test, even if both students are, in fact, equal in language ability. (b) Research has not established whether, or how much, partial knowledge of a system aids a person in remembering units of that system. Such research is required before we can infer that a student who knows a little English has a better chance of remembering English sentences than a student who does not know as much English. This is a reasonable assumption, but should be carefully investigated before it is used as the basis for proficiency tests. (c) Mimicry tests, such as the Del Rio and the MAT-SEA-CAL Oral Proficiency tests, require *exact* mimicry. Paraphrase is counted wrong (MAT-SEA-CAL Examiner's Handbook). Likewise, one section of the Northwestern Syntax Screening Test counts a paraphrase wrong if it has the effect of altering or eliminating the particular syntactic element that the test item intended to obtain. But if these tests intend to assess sentence processing or comprehension, there is no reason to score close paraphrases wrong. Students who make a small change in what is heard, leaving meaning intact, have probably understood the stimulus sentence. In fact, making a meaning-preserving change perhaps provides better evidence of understanding than exact mimicry. For example, a student who hears *He looked up the address* and repeats *He looked the*

**Tests employing this strategy are the Harris Articulation Test, the Ott Test of Oral Language, the Skoczylas Phonemic Unit Production Test, sections of the Milwaukee ESL Achievement Tests, the Language Assessment Scales, and the MAT-SEA-CAL Oral Proficiency Test.*

address up probably has not misplaced a word, but most likely applied the rules of English to produce a grammatical, synonymous version of the stimulus sentence. It is even possible that such paraphrase responses are more likely in the dominant language than in the secondary one, since students are more likely to control such synonymous, stylistic variants in the better-known language. Tests that require exact mimicry predict the reverse: in the dominant language, a person should produce *fewer* variations from exact, word-for-word repetition.

Passive Comprehension. In trying to determine whether a child understands a particular linguistic form or structure, a test can either attempt to get children to produce the structure in appropriate circumstances, or it can attempt to get them to act appropriately in response to hearing the structure. With the latter strategy, it is usually very difficult to be sure the children control the particular structure in question, even when responding appropriately. For example, in one item of the Language Assessment Scales, students must choose the correct picture in response to the passive sentence *The forks are held by both children.* Only one of the three pictures shows two children with forks, so it is possible to get this item right by simply combining the words *both* and *forks*. In any one case, we cannot be sure that a correct response to this item indicates control of the English passive construction.* If we wish to test for control of some specific structure or expression of English (or any other language), it is better to try to get the structure actually *produced*. It may well be unwarranted to infer that a particular structure is controlled merely because the response is appropriate.

Inferring Lack of Control from Lack of Performance. Some of the scoring schemes used on tests occasionally run afoul of the problem that lack of knowledge *cannot* be inferred from one or a few instances of lack of performance. In the Test of Grammatically Correct Spanish/English, for example, points are awarded according to the particular form students elect to use (e.g., person-marking, plurals, contractions, prepositions). Failure to use the scored items does not, however, indicate that they are not known or cannot be used by the students. A random factor is introduced into the final test score because results depend, in part, on what students happen to say on one occasion. Similarly, the Ott Test of Oral Language awards more points for production of a compound or complex sentence than a simple sentence, and more points for a simple sentence than a phrase. However, students might produce perfectly natural, correct answers to the test stimuli without ever producing a complex sentence. (Indeed, we know of no English question whose only natural answer is a full, complex sentence). It is not legitimate to conclude that students cannot produce complex sentences, when they have not been asked a question that naturally requires a complex sentence answer. While the El Paso Oral Language Dominance Measure is attractive as a dominance or proficiency test, it is somewhat less useful as a diagnostic tool. This test induces a variety of English syntactic and morphological structures (though the test materials do not specify exactly what structure each item is intended to elicit). In scoring responses to these items, the test avoids the erroneous scoring strategy previously criticized by allowing any grammatical and appropriate response as a correct answer. Although generally a good scoring strategy, it produces problems when we attempt to determine whether a child controls particular English structures, since appropriate answers are possible that do not contain the particular structures that the test intends to elicit. We cannot assume that failure to produce those structures means no control of them. Before the El Paso test could be used with confidence for diagnostic purposes, test items would have to be designed much more carefully to focus on specific linguistic structures or expressions.

**The test materials do not claim that this item assesses control of passive constructions; in fact they never specify what structure is tested by any item. However we may assume that a passive stimulus sentence is intended to test control of passive constructions.*

Self-Reported Data. The authors of the Spanish-English Dominance Assessment Test state explicitly that an assumption underlying the test is that children are able to report accurately on what language they speak in particular social situations (Manual). Under this assumption, it is justifiable simply to ask children what language they speak in given situations. Part of the Spanish-English Dominance Assessment Test is organized on this format, and a number of other tests we reviewed depend, wholly or in part, on the validity of self-reported data.* These are tests that we term "sociolinguistic questionnaires," which seek to determine language proficiency or (more often) language dominance by gathering sociological information about personal language use, rather than attempting to determine language ability directly through linguistic tasks. For a given language, these tests might ask: *Who do you speak the language to? Who speaks it to you? Do various relatives and friends speak the language? Where do you speak the language? Do you, your family, or your friends listen to music or radio, watch movies or TV, or read newspapers or magazines in the language?*

There are four problems with the sociolinguistic test format. First, tests that ask children whether they speak English or some other language in particular social situations assume that all the children will be consciously aware that they control two separate and distinct communication systems; but some young children who may be aware that they speak differently to different people and in different situations, may not be explicitly aware of the difference between the two systems, or of the names of the two systems. So, some young children might speak English in situations where it is socially appropriate to do so, but might not understand what is meant by such a question. (We will discuss this point further in connection with a different issue.)

Second, a person's honest beliefs about where, to whom, how much, and in what social situations a given language is spoken may not match actual practice. There is considerable sociolinguistic research literature that contradicts the assumption that children can accurately report their language use. These studies call into question the results of all the sociolinguistic tests we reviewed.

Third, children might knowingly give false answers to questions about their language, for a variety of reasons. Children might report what they think should be the case or what they want the examiner or school authorities to believe, rather than what is factual. This problem may tend to arise especially in situations where use of a given language is stigmatized (e.g., where it serves to identify a minority population of lower socioeconomic status in a given geographical area). In addition, there is always a possibility that older children, who may know that their answers will affect their placement in ESL or bilingual programs, may attempt to give answers that place them where they want to be. Two of the tests we reviewed, the Los Nietos School District Language Dominance Survey and the Moreno Quick Language Assessment Inventory, contain questions about language use which are to be answered by the students' parents. Obviously, the above criticisms do not wholly apply to these tests; however, it should be noted that parents' responses--for a variety of reasons--may reflect particular expectations, opinions, or biases which could interfere with accurate reporting.

Fourth, certain sociolinguistic questions might not even have a clear and unequivocal answer. How will children determine in what language their mother speaks to them, when she may speak different languages under different circumstances or tend to

**Tests we reviewed that make use of this format (wholly or partially) are the Brooks County Language Usage Inventory, the Hayward Language Dominance Indicator, the Language Dominance Criteria, the Language Dominance Survey, the Los Nietos Language Dominance Survey, the Marysville Test of Language Dominance, the Moreno Quick Language Assessment Inventory, the Robstown Oral Language Inventory, the Skoczylas Home Bilingual Usage Estimate, and the Spanish-English Dominance Assessment Test.*

code-switch (to speak both languages alternatively in the course of a single conversation or even within a single sentence)?

In connection with tests that include both a sociolinguistic section and a direct measure of linguistic performance, one further problem exists in the comparison of these two very different types of information. It is not at all clear how, or even whether, they can be compared. Some tests derive a total score for the two separate sections, and then simply add them to obtain an overall dominance score. This is something like adding apples and oranges, without some research on how the two types of questions are to be weighted. For example, in determining language dominance, the scoring procedure for the Brooks County Language Usage Inventory makes the implicit prediction that being able to identify a picture of a feather as a *pluma* counts exactly as much as claiming to use Spanish in the home. It is difficult to see any basis for this claim. This test and the Hayward Language Dominance Indicator both add the scores of sociolinguistic sections to those of direct performance sections. Two other tests, the Language Dominance Survey and the Language Dominance Criteria, are more careful in at least refusing to make any direct comparison of the results of sociolinguistic and language performance sections.

DO SPECIFIC TEST TASKS DEMONSTRATE OVERALL LANGUAGE SKILL?

A number of test tasks do measure what they intend to. It is relatively easy to construct a test that measures vocabulary knowledge or correct pronunciation, and very easy to construct a test that measures the number of words or sentences children utter in response to a given stimulus. However, the skills measured by these particular tasks may fail to provide a valid index of *overall* language ability.

The Value of Vocabulary Tests

The widespread appeal of vocabulary tests probably stems from the facts that they are easy to construct and score, and that they assess one of the most highly visible, straightforward aspects of language ability--knowledge of word meanings. Nevertheless, from a linguistic point of view, there are possible problems with vocabulary tests as measures of overall language skill, since there is much more to knowing a language than knowing words in that language. Vocabulary knowledge is highly detachable from other language skills. Memorizing a thousand common words of a language would probably enable students to do rather well on simple vocabulary tests, but more immediately central to proficiency is the ability to form normal, grammatical sentences and to use appropriate expressions under appropriate circumstances. By simple memorization, vocabulary knowledge in a second language can be increased without necessarily increasing proficiency in using the language for communication.

In defense of vocabulary tests, it could be argued that second language learners do not usually start out by memorizing a thousand common words; rather, they begin by learning a whole array of language skills--including vocabulary--and that a particular degree of vocabulary knowledge usually indicates a corresponding degree of proficiency in other language skills. Despite this probability, tests using vocabulary knowledge as an index of overall proficiency are of limited value. Vocabulary is more subject to variation than any other aspect of language. It varies among speakers of different ages, from different regions, and with different cultural backgrounds. It is also more subject to change over time.

Constructing vocabulary tests that account for all this variability is difficult. This is reflected by the fact that vocabulary tests almost always elicit *content* words (predominantly nouns, sometimes verbs, occasionally adjectives), and it is content words that are *most* subject to cultural, regional, and temporal variation. Vocabulary tests almost never elicit *function* words (prepositions, conjunctions, pronouns), which are more intimately associated with the grammar of a language. For example, the vocabulary section of the Pictorial Test of Bilingualism and Language

Development elicits nouns exclusively. Fifteen out of 18 items on one test of Level 3 of the Language Assessment Battery are nouns, and the remaining three are adjectives. In one vocabulary section of the James Language Dominance Test, 16 of 20 items are nouns, while four are present participles describing actions (*talking, eating*).

Based on the premise that if children do not know the common words of a language, they will not know the rarer ones, vocabulary test constructors often try to sample a "core" vocabulary of common words. To identify a core vocabulary, test makers sometimes depend on word frequency lists; the Language Assessment Scales, for example, makes use of the Thorndike and Lorge word frequency list, which was compiled in 1952 from a survey of children's reading materials. It is important to note that vocabulary tests that are based on such word frequency lists can only be as good as the lists themselves. As (if) the lists become dated, they may not accurately reflect children's current spoken language, which is what oral proficiency tests must assess. Any differences between how children spoke in 1952 and how they speak today would work against this kind of vocabulary test as an adequate measure of children's *present* oral English ability.

In addition, it is difficult, or even dangerous, to claim that any particular core constitutes an adequate vocabulary sampling for all children. Vocabulary is that part of language which--along with its other variables--is most subject to individual variation, most dependent on what experiences the individual happens to have had. Since the words a child knows will always depend, to some extent, on idiosyncrasies of the child's history, poor performance on a particular set of "core" words does not mean lack of general vocabulary knowledge (and thus overall language proficiency).

Some vocabulary tests do not claim to provide an overall measure of vocabulary knowledge, but merely to provide a *comparative* vocabulary measure for two languages. This is the claim made by the Spanish-English Dominance Assessment Test and the Pictorial Test of Bilingualism and Language Dominance. These tests elicit names of objects (or actions) in Spanish and English. If more Spanish than English words are produced, this is taken to indicate Spanish dominance, and vice versa. This format avoids the previous objection that if children do not know a word because they have never encountered the object it names, then that word counts neither for Spanish nor English dominance. This kind of test, however, can run into the problem that vocabulary is often culturally oriented. For example, a bilingual Mexican American adult reports that most of the Spanish words he knows are associated with cooking, while most of the words associated with cars, auto parts, etc., are English--reflecting details of his particular background. Unless cultural domain is taken into account, vocabulary tests run a risk of being seriously biased toward the language associated with the domain they represent. The Spanish-English Dominance Assessment Test contains a vocabulary test focusing on the kitchen and the yard; this would yield biased results for children who speak Spanish in the context of the home, but speak English away from home. Such children would naturally know more kitchen and yard words in Spanish, yet they might know more English words in other domains, and in fact be as proficient in speaking English as Spanish.

The Hayward Language Dominance Indicator allows more leeway, eliciting words associated with the kitchen, street, store, school, and doctor's office. This is a fairly wide range of domains, but the test authors do not claim to have checked whether bilingual children in their school district tend to speak English or Spanish in each of them. (There is very little explanation or rationale provided with the Hayward test. It is possible that the authors did attempt to balance domains in which Spanish is likely to be used with those in which English is more likely.) The Pictorial Test of Bilingualism and Language Dominance was apparently more carefully constructed. Using a test group of Mexican American migrant children in Texas, the authors included only items known in English *and* Spanish by at least 50 percent of the children, and for every item known by more children in Spanish than in English there was

an item known by more children in English than in Spanish. This seems to be a clever and effective way to balance a vocabulary test for cultural domains. However, this test would have to be reconstructed and recalibrated for use with other cultural (e.g., Cuban children in Miami) or language groups (e.g., Italian or Chinese children).

Our basic criticism of vocabulary tests as measures of language proficiency is that a child who is proficient in a language will certainly know many words, but a child who is not proficient may still know a good number of words. Simply knowing words does not guarantee a functional knowledge of the structure of a language and conventions for its use. Thus, if a vocabulary test errs, it will be more likely to err in overestimating the proficiency of a child who, for example, speaks English poorly. This is exactly the error one does not wish to make, because it has the effect of denying special educational help to some children who may need it.

The Value of Phonology or Pronunciation Tests

Phonology and pronunciation tests do not assess overall linguistic competence, and, as far as we know, no one has claimed that phonology/pronunciation tests provide adequate measures of overall language proficiency. Rather, these tests claim to assess only one component of oral language skill. Many tests are constructed fairly carefully on linguistic principles to test the ability to produce or distinguish phonological contrasts characteristic of a given language. Tests of English phonology are frequently designed to include English contrasts and sound combinations which do not occur in Spanish, and therefore are often difficult for Spanish speakers learning English. The Language Assessment Battery requires children to distinguish between *y* versus *j*, *sh* versus *ch*, *th* [ð] (as in *the*) versus *th* [θ] (as in *three*), among others. The Skoczylas Phonemic Unit Production Test grades children on pronunciation of *th* in *that*, *sh* in *she*, *ked* [kt] in *liked*, *ng* in *running*, *a* in *fat*, etc. Other tests (e.g., the Test of Grammatically Correct Spanish/English) elicit children's speech and grade it for something they vaguely term "correct" pronunciation. Apparently, this means how closely children sound like native speakers in the impression of the scorer.

If such tests measure only phonology or pronunciation and warrant no inferences about overall language skill, then their value depends on whether or not it is useful or important to assess children's phonological knowledge or pronunciation skills per se, and whether test instruments are required to make this assessment. For educational purposes, it is certainly necessary to determine whether students' pronunciation of English allows them to be understood. However, any teacher who speaks English can make this determination without employing a sophisticated phonology test. Such tests may provide more detailed analyses in assessing pronunciation of certain sounds, but there is some doubt whether this kind of analysis is really useful in assessing English ability, since any *one* mispronunciation is not very likely to impede communication. A child who says *A chip with tall masts floated in on the tide* will not be misunderstood. It is only when mispronunciation becomes global and severe that communication is threatened, and a child's ability to get along in an English-speaking classroom is placed in jeopardy.

The detailed phonological diagnosis some tests make of students' speech production or perception might provide information that a teacher can use to help students with specific pronunciation problems. If the test shows that a Vietnamese child does not produce word-final consonants in English speech, then the teacher can focus on that particular problem. The Ott Test of Oral Language explicitly claims to provide such diagnostic input to English teaching. This sort of information might indeed be one useful feature of phonology/pronunciation tests, but past a certain point, there is a question of how much instructional time one would wish to spend on phonology. If pronunciation is understandable, then there are probably more important aspects of language to work on. It will be difficult, time consuming, frustrating, and probably unnecessary for students to try to produce native-sounding pronunciation.

It is certainly necessary to determine whether pronunciation can be understood; however, it is doubtful whether the detailed analysis provided by these kinds of tests is necessary to make this gross judgment. Furthermore, it is doubtful whether one would want to use the detailed analysis in making finer-grained judgments about linguistic ability. Detailed analysis might help teachers focus on a particular or acute pronunciation problem, but this, too, might be a mixed blessing if it overemphasizes pronunciation and neglects other aspects of language.

Measures of Linguistic Creativity

Some test items require students to exercise some degree of linguistic imagination by describing something or answering open-ended questions. This is not entirely unfair on language proficiency tests, since these demands are reasonably natural linguistic skills which students might have to use outside the testing situation. But, all such items--to some degree--measure a skill beyond simple language knowledge in that they also test creativity, imagination, and willingness to extrapolate from known facts, in the context of the testing situation. Whether these linguistic "creativity" items really assess language proficiency, or go beyond it, depends on how much extrapolation is demanded by the item and on how the results are scored.

Before we discuss particular tests that make greater or lesser demands on a child's imagination, we will present one argument for why it might be dangerous to place too heavy an emphasis on linguistic creativity as a measure of language ability. There is well-known (if anecdotal) evidence for two distinct types of second language learners: *Risk-takers* and *conservatives*. Risk-takers are readily willing to speak the second language and to make any number of errors, as long as they can make themselves understood. Conservatives, on the other hand, are much less ready to speak in the second language unless they are reasonably sure that what they are going to say will be grammatically correct. Furthermore--and most important--this distinction does not necessarily correlate with how well a person speaks the second language. Conservative language learners may be relatively proficient in the language, and still be unwilling to speak as freely or to say as much as a risk-taker who may be less proficient. Without claiming that every person falls in one of these categories, it is clear that if the distinction is at all real, it may affect the results of any test task that requires language production; but, it seems likely that it will have more effect on tasks involving linguistic creativity. A risk-taker is more likely to be willing to respond to such a task, and will probably produce more speech, and extrapolate farther from simple descriptions of known facts, than a conservative with the same degree of language knowledge. If performance is scored according to the grammatical correctness of the response, then this would tend to favor the conservatives. If it is scored according to the amount of speech produced or the degree of imagination displayed, then it favors the risk-taker. Either way, the creativity task makes it possible to confuse language ability with how willing a person is to take risks in speaking a second language.

To confound matters even more, systematic cultural differences may have an effect on this task. For cultural reasons having nothing to do with relative language proficiency, children from some cultural or national backgrounds may be reluctant to respond at length, while others from different backgrounds may be more verbally assertive in the same situation. Additionally, cultural background and/or personality of the test administrator may tend to encourage or inhibit children taking this sort of test. Thus, cultural differences and differences in individual style can play a confounding role in tasks attempting to measure language creativity or imagination.

Tests that we reviewed required linguistic creativity or imagination to varying degrees. Some tests merely contained a few items intended to make a child extrapolate from known facts or pictures; at the extreme, some tests were entirely composed of questions intended to extract connected stories which were then graded for

creativity, organization, "abstractness," etc. An example of the first type is the "Picture Generating Response" subtest of the Caso Test for Non-English Speaking Students. One series of pictures shows saws and boards, and ends with a picture of a hand holding a board. One question asks, *Who do you think the person may be in [this picture]?* and *Why do you think it should go in [this picture]?* It is not clear from the test materials how such items are to be scored--for the grammaticality of the answer, its content, or both. The Ilyin Oral Interview contains some similar items, and similar puzzles for the scorer.

A test that relies much more heavily on linguistic creativity is exemplified by one part of the Pictorial Test of Bilingualism and Language Dominance. A picture is shown to a child who is asked to tell a story about it. After the child responds with a story (or if he fails to do so), four more explicit probe questions are asked about the picture. This procedure is followed for English and Spanish, with the same pictures and probe questions being used for both languages. The child's responses are then scored for grammaticality and "completeness." Grammatical, complete sentences earn the best score; ungrammaticalities and phrases are next, and word lists naming the objects in the picture are scored lowest. This is not entirely a creativity test, because grammaticality and other formal properties of the child's answers are considered in scoring. Performance, and thus score, can be affected by whether the child is interested enough by the picture to think up a story about it--a factor not directly related to language proficiency. This "interest" factor will become much more important in the next two tests discussed, where straightforward description earns a lower score than flights of creative fancy, and the grammaticality of the child's response is apparently not seriously considered in judging proficiency.

In one section of the Oral Language Evaluation, the examiner shows the child a picture, reads a brief description of it, and then instructs the child to *tell about the picture*. Performance is scored according to six categories which represent the authors' beliefs about how language develops (Teacher's Manual), but which do not appear to reflect the findings of any recognized research in first or second language acquisition. In effect, the test seems intended to elicit a short story, since a simple description of the picture earns a much lower score than an abstract story containing details of mood and characters' emotional responses (Teacher's Manual). A straightforward description is taken to mean that a child "needs some aspect of language training [and] is probably not ready for reading." As this test is designed, such conclusions are entirely unwarranted. There may be many reasons why a child might not produce an abstract story. One could be that it might be very unclear to the child what is being asked when subsequently instructed to "tell about the picture" which has already been described by the examiner. There is no effort to explain what kind of response is desired, nor could there be: What could the child make of such words as *mood* and *emotional reactions*? So, it seems unlikely that the child will even be able to clearly understand the task posed.

This section of the Oral Language Evaluation allegedly assesses the need for oral language training and reading readiness; however, the test really confuses proficient language performance with a certain derivative language-based skill--the ability to compose fiction. Some people are good at composing fiction and some are not, but this does not necessarily say anything about their ability to use English under ordinary circumstances.

Most of these criticisms of the Oral Language Evaluation also apply to the Language Facility Test. Both tests set up a series of categories as a supposed hierarchy of language development, unrelated to any of the considerable body of research on language development. In both tests, the lower categories are related to the form of the language used, the linguistic elements that occur, and the organization among them; the higher categories contain judgments about the completeness, implications, organization, or abstractness of the story. The lower categories seem to evaluate

something very different from the higher ones, because judgments at the lower levels are based on the form of the language used, while the higher levels depend on the literary properties of the story the child tells. In scoring the Language Facility Test, clear instructions are given to ignore vocabulary, grammar, and pronunciation. A high scoring story could be "grossly ungrammatical and from a person not familiar with the correct names for some elements of the pictures" (Test Administrator's Manual). This scoring strategy makes us wonder just what is being tested here, since a person might not control enough English to function in daily life, and yet be rated in the very top category of this test.

Counting Linguistic Elements

Some tests feature a scoring strategy wherein speech--anything from lists of words to connected descriptions or discourses--is elicited from a child, and linguistic elements contained in this speech sample are counted. Whether sentences, phrases, words, or syllables are counted, and whether the final score is represented as a simple total, or as an average of one sort of element with respect to another (e.g., average number of words per sentence), the intent of such strategies is usually to claim that the inclusion of more elements indicates greater proficiency, and that the quantity of language produced is a good index of overall ability.

The general problem with this claim is that the total amount of speech a child produces on one occasion, in response to a particular set of stimuli, does not necessarily tell us what the child knows about language or is capable of producing. There are various reasons why children might not, on particular occasions, produce a lot of speech: (a) they might not know much English (or whatever language is being tested); (b) they might not be interested or motivated by the test items to speak; (c) they might not be linguistic risk-takers; (d) they might not have understood that they were expected to produce an extensive amount of speech; and/or (e) they may think that simply answering or responding concisely and to the point is all that is required. Tests that score performance by counting linguistic elements tend to assume that the only possible reason for children not to say much is because they do not know much of the language, but the other reasons are real possibilities and tend to vitiate the results of such tests.

The Basic Inventory of Natural Language (BINL) is scored almost entirely by counting linguistic elements the child produces. Various sorts of elements are counted and reported as independent categories in each child's "oral language profile." We will consider in some detail the two categories of "average level of complexity" and "fluency."

In the BINL, the average complexity level is found by counting complete and partial sentences, modifiers, phrases, and clauses, calculating a weighted total of all these elements, and dividing by the total number of sentences. This score is supposed to reflect the average complexity of the sentences the child produced, but there are several problems with this strategy. First, there are no accessible, objective, and reliable criteria for deciding what a clause is, what a phrase is, or what constitutes "modification." This sort of decision may appear simple, especially concerning the most common, clear-cut cases in each category, but everyday language makes constant use of a wide range of grammatical structures which challenge analysis. This fact is bound to complicate the scoring of a test such as the BINL and to limit its reliability. The difficulty with scoring on the BINL is further increased by the problem of ellipsis, although the authors of the test intend for appropriate ellipsis to be allowed (Instructions Manual). For example, in a question and answer sequence such as

Where is the cat in this picture?

In the tree.

the answer would be intended to count as a complete sentence, and not as a prepositional phrase. Looking only at the language produced by the child being tested,

the scorer of the BINL has no way to distinguish acceptably elliptical utterances such as the above, which are to be counted as "complete sentences," from "partial sentences" or "phrases."

Another problem with this scoring strategy is that it assumes that all linguistic phrases, all clauses, and all forms of modification are equally complex. A moment's reflection will show that this is not true. The scoring procedure is apparently not even intended to consider whether a child's responses are appropriate, internally related, or even grammatical. The test materials do not say that only grammatical sentences should be counted as "complete sentences." In fact, certain ungrammaticalities are acceptable (Instructions Manual), but it is not clear whether the authors intend to accept all ungrammaticalities. A more subtle problem concerns the coherence or relatedness of the response, because a series of unrelated sentences would be scored as high as a coherent response. For example, the following two responses would be scored as approximately equivalent performance, but the first is obviously a more sensible use of English:

John went to the store. He bought some bread. Then he came home.

The cat is on the mat. My ear is purple. Beavers live in Canada.

Another category in the BINL language profile is the "fluency" score--the total number of words produced. Here a child can get a higher score for responses that are over-explicit to a degree that is linguistically unnatural. In the Instructions Manual, the test authors point out that the natural answer to the question *Where are the toys?* is *On the table*, not *The toys are on the table.* In fact, in most circumstances it would be unnatural--uncharacteristic of the ordinary conversation of native English speakers--to produce the complete sentence answer. Yet, children will get a higher "fluency" score on the BINL for producing this conversationally unnatural response.

The Oral Language Dominance Measure (El Paso Public Schools) assesses English or Spanish proficiency as a function of two variables--total number of grammatical and appropriate responses to the test items and total number of words produced. The first of these is linguistically sensible as a measure of proficiency, but the second seems dubious. The Examiner's Manual reports that the test makers' own research has found the two factors to predict language ability, but this research is not described, and no bibliographical references are given.

The Test of Grammatically Correct English/Spanish (Las Cruces School District) scores responses, in part, by counting the number of words and syllables produced, and by evaluating the uncommonness of the words used (i.e., whether they appear on a list of the statistically most common English words). Thus, this test partly evaluates responses according to whether long words and/or uncommon words are used. In effect, this scoring procedure focuses on the content words used by the child, since the function words of English are characteristically short and frequent in occurrence.

The Language Dominance Indicator Test (Hayward Unified School District) contains two parts. One part asks five questions about the use of English or another language (*When your father speaks to you, what language does he use?*), and one point is scored for English each time the child claims to use it in a particular social situation. Five English points are possible. The other part of the test is scored by counting the number of nouns produced; the child is asked to name objects found in the kitchen and in four other locations. One point is scored for English for every object named in English. Aside from the problems involved in measuring language proficiency by simply counting words, this scoring scheme implicitly makes the unsupported and counter-intuitive claim: determining if children know the English word *stove* as something found in the kitchen tells us exactly as much about their English proficiency as determining that the father speaks in English. While it is fairly clear that these are not equivalent indices of language ability, we have no idea how they should be compared, nor whether they should even be compared.

LINGUISTICALLY ARTIFICIAL TEST TASKS

Generally speaking, a test of language proficiency should assess familiarity or facility with the rules of a language and the conventions that govern its use as a means of communication. In testing language proficiency, we should try to avoid assigning requirements beyond those imposed by the ordinary conventional use of language. Such extra requirements--artifacts of the testing situation--risk getting in the way of what the test is really trying to measure--overall language skill. Of course, all tests will be somewhat artificial. The testing situation is not just a normal conversational interchange, and few, if any, children will fail to realize that they are being tested. But, the language of tests and various aspects of the tasks imposed should match ordinary language usage as closely as possible in order to avoid imposing linguistically artificial tasks on the child which require extraneous knowledge. Such knowledge might indeed be language-related, but goes beyond what would properly characterize language proficiency.

Understanding the Test Task

Occasionally the instructions telling the child what is required by particular test tasks are excessively vague about what kind of response will constitute good performance. This becomes a problem in cases where something more is required of the person taking the test than simply observing the rules and conventions of normal English usage. Children who manage to figure out what is required may do better than those who do not, although they actually may not be more proficient at English (or whatever language is being tested).

The most common case involves test items that intend to elicit a particular linguistic structure, but which in fact will not do so unless a complete sentence response is produced. Since the rules of English do not very often require complete sentence responses, tests must resort to extra instructions or other manipulation in order to elicit complete sentences. Many tests simply instruct the child to *answer in complete sentences*; other tests employ more indirect strategies, using probing techniques or teasing to elicit the target structure from the child. Because the intent of the probes is often not clear, these methods can be very puzzling. The Ilyin Oral Interview states, "It is important that the respondent give complete sentences, both statements and questions. If he gives short answers, the interviewer should instruct him to give a longer one." Quite apart from whether it is natural to demand complete sentence answers to all questions, how does the child know that "longer" means "complete sentence"? Similarly, in the directions for the Moreno Oral English Proficiency Placement Test, examples are given for eliciting the required complete sentence responses. In reply to a test question, *Does Juan have a cat?*, a child might well answer *No*. This is a perfectly natural English answer, but since it does not serve the test purposes, the examiner is instructed to prompt, *No, what?* If the child answers [a] *policeman* in response to *What is he?*, the examiner is instructed to prompt, *Tell me that...* or *Tell me more*. It is difficult to predict what children might make of these questions, and it may be difficult for them to interpret what is expected. Assuming the answer to the original test question is truthful, grammatical, and appropriate (as is the case with the two Moreno examples), it might be confusing to be probed for a further answer--why is the examiner not satisfied? The child might even think a mistake has been made, when in fact the given answer is correct and natural, but not the one that the test question was designed to induce.

This case and others involve test instructions that may give test-wise students an advantage over children who are inexperienced at taking tests. In these tests, there is an implicit test-taking skill--quite apart from purely linguistic abilities--which can increase the possibility of a high score. So, individual experience may play an important role in determining scores.

The Crane Oral Dominance Test consists of a memory task which presents children with eight words, four in English and four in Spanish. Children must recall as many of the

eight as possible, in either language; if more English words are repeated, then the child is judged English dominant and likewise for Spanish. Children are told to *say [the words] any way you want, in English or in Spanish* (Philosophy and Directions Booklet). Nevertheless, some children may perceive it as part of the test task to remember what language each word was presented in, and might make a conscious effort to produce four English words and four Spanish words. To the extent that they are successful at this, they will tend to be judged as balanced bilingual--dominant in neither language. Other children may simply try to recall words in either language as the authors intended, and their scores will be free of the influence of a conscious effort to equalize the number of words. Thus, if all children do not perceive the test task the same, it may be unsound to compare the final scores of any two and claim that the test shows one student to be more English- or Spanish-dominant than the other.

Sometimes particular items or sections of tests require instructions that pose a greater linguistic challenge than the actual response required. In this case, the test runs the risk of concluding that students do not know the correct response when, in fact, they may be having difficulty understanding the more complex instructions. For example, on the Listening Comprehension section of the Shutt Primary Language Indicator Test, each item is intended to assess comprehension of a noun phrase. Some items test simple noun phrases (*a banana*); most items test nouns with relative clauses (*a child who is pitching, something tied with a string*); some test relative clauses that lack preceding nouns (*what a child should take to the teacher on the first day of school*). In each case, however, the noun phrase being tested is embedded inside an expression that is grammatically quite complex: *Make an X in the square on the picture which shows* ___. For most of the items on the Listening Comprehension subtest, this expression is more complex than the noun phrase expression being tested; this adds considerable irrelevant difficulty to the task, making it very hard to pinpoint exactly what is not known by students who miss such an item.

Ellipsis and Complete Sentences

Many tests provide a definite instruction to *answer in complete sentences* or resort to some other strategy to try to elicit complete sentences where they would not normally be required in conversation. We criticized some of these covert strategies earlier on the grounds that they may be confusing or misleading. However, what is wrong with straightforwardly telling children to respond in complete sentences? Just this: knowing the correct form and use of complete English sentences (indeed a component of English proficiency) is not the same as knowing the meaning of the expression "complete sentence."

It has been demonstrated that people who can speak English proficiently do--in some sense--know what constitutes a complete English sentence. They can construct and use complete sentences when the conventions of English usage require it, and when they omit complete sentences in their speech, they usually do so in certain principled and specifiable ways. Native English speakers who do not know that a complete sentence consists of a subject, predicate, etc. still know how to use complete sentences when it is appropriate. The ability to use English in this way is knowledge *of* language; the ability to construct and produce complete sentences on command is knowledge *about* language. Individuals with the skills to produce sentences on command generally have acquired them by formally studying and learning about English grammar, but lack of formal training does not necessarily indicate a lack of English ability. Children who fail to respond in complete sentences when explicitly told to do so might not have acquired the skills to construct complete sentences, or they simply might not know what the term "complete sentence" means. It is possible that some measures claiming to test English ability are testing explicit knowledge of the concepts of English grammar.

Previously, we criticized strategies used in the Ilyin Oral Interview and the Moreno Oral English Placement Test for trying indirectly to obtain complete sentence

responses. Tests explicitly instructing children to produce complete sentences are the Short Test of Linguistic Skills, the Caso Test for Non-English Speaking Students, and the English as a Second Language Achievement Tests, Level 1. The Short Test of Linguistic Skills tells the child, *I am going to ask you some questions in English, and I'd like you to answer the questions by using a complete sentence* (Examiner's Manual, emphasis theirs); the others are equally clear on this point.

The pervasiveness of ellipsis in normal language usage creates problems with the scoring strategies used on several tests. On one part of the Ott Test of Oral Language, students answer questions put by the examiner. Although not directly instructed to use complete sentences or prodded to do so, students' responses are graded according to a hierarchy in which simple noun phrases get one point, verbal and other phrases get two points, and complete simple sentences get three points. This scoring hierarchy is applied to the answers for all test questions, even though conventional English usage requires different kinds of answers to different kinds of questions. Questions like *What do you see here?* invite noun phrase answers (*a dog, a man in a red hat*) as the most natural, straightforward replies, and questions like *What are they doing?* invite verb phrase answers (*playing football*). The scoring scheme for the Ott test would award a better score for unnatural, overly explicit answers to the above two *what* question examples. The short answers that these questions would ordinarily evoke would earn children a lower score. In effect, this section of the Ott test rewards complete sentence answers without regard to whether such answers are natural, and without informing children what is expected. If a scoring system rewards performance which deviates from standard usage, test subjects must be told what kind of performance is desired.

The scoring system for one section of the Pictorial Test of Bilingualism and Language Dominance likewise rewards the use of complete sentences. Like the Ott test, the Pictorial test does not specifically elicit complete sentence responses, but still assesses a higher degree of proficiency in English if such responses are produced.

The scoring scheme for the Basic Inventory of Natural Language (BINL) also penalizes partial sentences, but ellipsis is nominally allowed. Correctly used, elliptical sentences are to be scored as if they were complete sentences (Instructions Manual). Although an acceptable scoring procedure, there may be some problem in determining whether an elliptical response has been correctly produced. The problem could arise due to BINL data collection procedures, where data is collected by gathering children together in a group and having one child describe or tell a story about a picture to the other children. Conversational interchange among the children is encouraged on the assumption that it will serve to elicit natural language from the story-telling child. Albeit that the assumption is probably correct and the situation ideal to elicit instances of ellipsis, the data that is transcribed and subsequently scored apparently consists only of the speech of the particular child telling the story. Comments and questions from other children in the group are lost by the time the speech sample is scored. (The sample may be mailed away for scoring by the test publishers, who obviously never heard the group conversation that served as the context for the speech sample.) In order to determine whether an elliptical form has been used correctly, it is crucial to know what else has been said in the context where the form occurred.

The Michigan Oral Language Productive Test provides a series of questions intended to elicit particular grammatical forms, particular words, and particular pronunciations. Since correct elliptical answers to the test questions would frequently not include the structure being tested, the Michigan test adopts a strategy in which the examiner prompts the child with the first word or two of the desired response. For example, in one item intended to elicit past tense, the examiner points to a picture of a fish and says, *Where did the fish jump? The fish....* The child must say *The fish jumped in the river*, using the past tense verb, *jumped*. This strategy is necessary to prevent the child from answering *in the river*, which is a perfectly natural

answer, but one that does not contain the desired past tense verb. This strategy, as with others we have discussed, makes the test task very artificial. In ordinary conversation when asking a question, we do not prompt the first few words of the expected answer; when answering a question, we do not expect the questioner to provide the first few words of the desired response. It is difficult to make test tasks approximate ordinary language usage, but tasks like this one run the risk of testing understanding of and willingness to comply with the artificial test task, rather than the ability to handle ordinary English usage. The more variation there is between the test task and ordinary usage, the more room there is for doubt that the test actually assesses ordinary skills.

In some cases, the prompting strategy employed by the Michigan test will not necessarily be successful in eliciting the desired form. For some items, there may be grammatical, coherent, and factually correct responses, repeating the examiner's prompt, which still do not include the desired grammatical or phonological feature. In one item, the examiner says, *The father wears shoes in this picture. Tell me if the boy wears shoes. No, he....* The child must answer *No, he doesn't (wear shoes)*, because the item intends to test negated *do*; but the child might equally say *No, he has no shoes* or *No, he's barefoot*. Are these answers to be counted wrong? The MAT-SEA-CAL Oral Proficiency Test employs the same prompting strategy just described for the Michigan test, and is subject to the same criticisms.

In conclusion, we will mention a few tests that are noteworthy in that they do *not* require complete sentence responses, nor do they score elliptical sentences lower in environments that naturally permit ellipsis. The Oral Language Dominance Measure states that "any grammatically correct response in the language being elicited is considered acceptable" (Examiner's Manual), whether or not it contains the specific structure that the item was designed to induce. Examples of permissible responses make it clear that ellipsis is considered acceptable. Similarly, the speaking tests of the Language Assessment Battery allow for ellipsis, and other correct and appropriate answers that differ from what the test item was intended to elicit (see Examiner's Directions for Administering, Level 2). The Bilingual Syntax Measure also allows elliptical and other appropriate, grammatical responses as correct answers to test items (Manual; Technical Handbook).

Metalinguistic Instructions

The requirement to produce complete sentence responses poses a metalinguistic task for children in that it requires not just knowledge of a language but also some explicit, conscious knowledge about a language, about the analysis of its grammar, and about the names grammarians have invented to refer to elements of the language. The complete sentence requirement is by far the most common metalinguistic task posed by language proficiency tests, but some other tasks and instructions found in the tests we reviewed also require metalinguistic knowledge of one sort or another. A metalinguistic task creates the possibility of confounding--test items may not be measuring the linguistic skill they intend to. If by instructing them to *ask a question*, a test seeks to determine whether children can construct an ordinary English question, children's failure to do this properly could be caused by ignorance of *how* to construct an English question (as the test predicts), ignorance of the meaning of the instruction "ask a question," or ignorance of the word "question." As with complete sentences, children may be able to construct and use questions properly and effectively, but not be able to produce one on cue.

The Ilyin Oral Interview contains a number of items of this sort: *Please ask me a question about this picture. Please ask a question about this picture using the word "after."* Several of these items are followed by the instruction, *Please answer your question.* Another item says: *This picture shows Bill every weekday. Please ask me a question about this picture.* Here, the desired response is presumably something like *What does Bill do every weekday?* Apparently, these items aim to control the

form of a response by explicitly indicating what form to use and--in some cases--what words to use. One result of this is to produce a conversational interchange that would never be found anywhere but on a language proficiency test. We usually ask questions because we want to know something, not because someone instructed us to ask a question. People sometimes require us to say certain things, but they seldom tell us what words to use in saying them, and we are very seldom instructed to answer our own questions. Performing such artificial tasks requires certain test-taking abilities, as well as the ability to speak English; yet, it is only the latter ability that is being tested. If the task is very unlike the usual application of the skill being tested, then test-taking skills become relatively more important, and we are less likely to be measuring what we want--proficiency in the ordinary use of English.

A related metalinguistic problem is found on the Harris Articulation Test. This test provides a detailed assessment of children's pronunciation of English consonants in initial, medial, and final positions in words. Pictures are presented to children, who must name the objects. For example, a picture of a ball is presented to elicit the word *ball* so that pronunciation of initial *b* can be judged. To get consonants in initial position, a single noun must be produced without an indefinite article preceding it. If the child says *a ball* instead of *ball*, then an initial position *b* has not been produced, and the test does not obtain the information it seeks. In an effort to control this, the examiner instructs the child who produces an indefinite article to *only say the name of the picture*. This may be a puzzling instruction. Children who identify the picture as *a ball* may think they have already said only the name of the picture and might wonder why that response is not acceptable. This strategy is reminiscent of tests that try to probe for a complete sentence response or, failing to get a complete sentence, ask for a longer answer.

Another example of a metalinguistic instruction involves tests that explicitly tell children to respond in a particular language. In language proficiency testing, it is legitimate to expect a child to be able to respond in the language used to administer the test. Many tests take advantage of this by simply presenting test items (oral questions or commands) in the language that is being assessed. For example, the Spanish/English Language Performance Screening clearly instructs examiners to use only Spanish in the Spanish section and only English in that section. The Oral Language Dominance Measure separates the two languages completely by requiring a lapse of several days between administration of the English and Spanish sections. These are logical and probably effective strategies for eliciting responses in a particular language. In contrast, it is not necessarily reasonable to expect that all children will be able to respond in the appropriate language when instructed to do so. Children might know both languages and be able to separate them (in the sense of being able to speak the proper language in the proper social situation), but still not understand the metalinguistic instruction to "answer in English." This is unlikely to be a problem with older bilingual children, who are probably well aware that they speak two languages, but some younger children may be aware of speaking in a certain way with some people and differently with others without being overtly aware of the distinction represented by different language names.

The Pictorial Test of Bilingualism and Language Dominance introduces a picture-naming vocabulary test with the instruction, *You may answer in either English or Spanish* (Manual). The Crane Oral Dominance Test, as previously described, sets a memory task in which the child must recall a set of orally presented words. Instructions include "Sometimes I will say the word in English and sometimes in Spanish. Listen carefully. When I finish, name as many words as you can remember. You can say them any way you want, in English or Spanish" (Philosophy and Directions Booklet). Understanding this instruction is crucial to the Crane test, since a natural response might be to report the Spanish words in Spanish and the English words in English, which would seriously interfere with the test's intent.

The James Language Dominance Test has perception and production sections in Spanish, and then in English. The only instruction that intervenes between the Spanish and

English sections, to inform the child that a change has taken place, is the examiner's statement, *We're going to talk in English. Look at these pictures* (Manual). The abruptness of the change from Spanish to English might pose some problems in shifting from one language to the other. (Contrast the complete separation of languages required by the Oral Language Dominance Measure.) The James test also recommends using a metalinguistic instruction if the child responds in the wrong language. For example, on the Spanish section, if the child responds in English, the examiner is instructed to say *Dímelo en Español*. As we indicate above, this may not be an effective way to obtain Spanish responses from some young children.

Finally, the Flexibility Test to Measure Language Dominance in Spanish-English Bilinguals is a test of a unique type, in which children presented with a set of eight scrambled letters are first told to construct as many Spanish words from those letters as possible in 60 seconds, and then to construct as many English words as possible from the letters. This process is repeated five times with five different sets of eight scrambled letters. To accomplish the task, children must be able to make a clear and conscious distinction between English and Spanish when instructed to do so. So, this task is open to the prior criticisms that successful performance requires a certain kind of metalinguistic knowledge.

Identical Test Items in Two Languages

Another way in which the overall organization of test tasks can depart from everyday language usage is by requiring students to answer exactly the same questions twice--once in each language. Although this is done so that competence in the two languages can be directly compared, this kind of task is unlike any conversational use of language and even unlike the way language is customarily used in other types of tests. Ordinarily, there is no point in asking the same person the same question twice. Either the question is answered correctly or it is not, and for testing purposes, that is all that is required.

The identical item format usually asks simple short-answer questions, which decreases the pretense that the testing situation is just an ordinary conversational interaction. It is open to argument whether it is worthwhile to try to maintain any such pretense. Some tests try to do so overtly (e.g., the Basic Inventory of Natural Language and the Bilingual Syntax Measure); many more do not. Tests should assess ordinary conversational language abili and normal language usage as much as possible, and tests that adopt features different from normal usage must show that they do not interfere with the assessment of the ability to use language in ordinary circumstances.

Even though it is conversationally unnatural to do so, a bilingual child will probably have no trouble answering a question like *What is your name?* in both languages. However, tests which require longer, descriptive, or creative answers, and which duplicate the task in two languages, may indeed create bias problems. One part of the Pictorial Test of Bilingualism and Language Dominance involves showing children a picture and asking for a story about it--first in English, then in Spanish. The children, of course, have the option of making up two different stories about the same picture, but if they choose to tell the same story twice, it is likely to be shorter and less elaborate the second time. Children may not feel it necessary to be as careful or precise about introducing characters just introduced or describing situations already described. They may use more pronouns and more ellipsis the second time, assuming (correctly) that the examiner will understand what is being said, since it all has been said before. The test, however, is scored according to the completeness of the thoughts expressed, and the demonstrated expression of complete subject-verb-object sentences (Manual). Children's performance in the second language is liable to appear less complete, and so less proficient, purely as an artifact of the test format.

One section of the Spanish-English Dominance Assessment Test also adopts the format of picture description. Examiners create their own pictures representing local conditions, but it is not clear from the test materials whether children are required to describe the same picture twice, in Spanish and then in English, or to describe one picture in Spanish and a different one in English. In view of the above considerations, it would seem preferable to follow the latter procedure.

Finally, one way to overcome the problems with identical test items in two different languages is to administer the English and non-English sections of the test on different occasions. After a lapse of a few days, it might be less unnatural to ask children to answer questions already answered. The Oral Language Dominance Measure follows this procedure.

IMPLICATIONS OF SECOND LANGUAGE ACQUISITION FOR TESTS

Most tests which purport to assess relative English proficiency are implicitly based on an idea that people tend to learn structures which are easier, sentence types which are simpler, and vocabulary which is somehow more basic, before mastering more complex and difficult parts of the system. While this is a reasonable assumption, the content of many tests still is based on notions of complexity which are supported neither by linguistic theory nor by empirical evidence of how children acquire English as a second language.

Children acquire language through a systematic developmental process in which they actively and creatively build up knowledge of the language system. Through contact with language in everyday situations, they gain a sense of how utterances in the language are structured and organized, and gradually they integrate these linguistic structures into an internalized grammatical system for the language. This process is frequently termed language *acquisition* to distinguish it from conscious, deliberate language *learning*.

There are various kinds of evidence which suggest that the language acquisition process is guided by some property of cognitive organization--some special capacity for dealing with language structure which all children have. Among the kinds of evidence advanced for this view are the following:

- Children learning any language not only repeat and imitate speech, but they also produce new, unique sentences which they have not heard before.
- Many of the "errors" made by children learning English as a second language are similar, regardless of the native language of the child.
- Second language learners tend not to transfer syntactic patterns of their native language to the new language; errors appear to be developmental rather than to result from first language interference.
- Stages in the language acquisition process appear similar in children from different learning environments.
- Stages in the second language acquisition process are apparently not guided by external input factors such as the auditory salience of speech elements or corrective feedback from other people (although this point is somewhat controversial).

The acquisition of language structure, then, tends to be a systematic and regular process. As they build competence in the new language, all children tend first to master certain types of structures, later developing control of somewhat more complex structures. In the acquisition of English as a second language, characteristic patterns of word order for English sentences are typically mastered before any higher-level markers of grammatical relationships. Research also shows that use of the singular copula *is* is controlled before the short plural marker *-s*, and that both

of those structures are controlled before the possessive morpheme *'s*.* Other research has shown that children learning English as a second language use *is* before they use *do* and *can*; furthermore, children use all three of these verb forms before beginning to inflect any of them for number or tense (as in *was, are, did*), and also before other modals (*will, would, must*) appear in children's speech. These are just a few examples of regular patterns which all children acquiring English as a second language tend to follow.

The acquisition of English grammatical morphemes is not a linear and additive sequence; rather, structures seem to fall into groups, and children tend to acquire the structures within one group before they gain control of those in another group. This hierarchy of English structures--which is supported by empirical evidence from language acquisition research--predicts that if children control some grammatical structure which is relatively high in the hierarchy (i.e., which is predicted to be a relatively late acquisition), they will also control structures which are acquired earlier. Conversely, children whose speech does not indicate control of some relatively low-level structures (i.e., relatively early acquisitions) will probably also lack more difficult structures.

The fact that there are such hierarchical patterns in second language acquisition has important implications for testing English proficiency. These patterns make it possible--in principle at least--to design a test in such a way that when children demonstrate control (or lack of it) of certain grammatical structures, we can learn more about their language ability than just their facility with those particular structures. We can also infer something about placement on the putative language acquisition hierarchy, and thus gain information about the developmental level in the acquisition of English.

Type of Language Knowledge Tested: Discrete Point vs. Integrative Testing

For some children there may be a gap between the ability to use English in natural interaction, and the ability to perform formal tasks in situations which invite conscious attention to language form (such as some testing situations). The first type of ability will reflect a stage in the developmental process of language acquisition; the second will not.

In evaluating tests which claim to assess students' control of English grammatical structure, it is important to recognize which of these two types of language knowledge a test is trying to assess and what the test reveals about students' developmental level of English acquisition. Depending on the design of the test, it may elicit representative evidence of students' internalized control of English structure, or it may yield an inventory of some of the precise rules of grammar students have learned. In the former case, a test score might indicate (in part) students' ability to use English in meaningful interaction, and placement in the developmental process of acquiring English. In the latter case, the test score might provide no information about either of these important points.

English proficiency tests which include some measure of grammatical ability usually fall into one of two general categories: *integrative* and *discrete point*. Integrative tests obtain evidence of students' overall control of the language by having them produce connected discourse in some meaningful context. For example, the Basic Inventory of Natural Language, the Language Dominance Survey, the Language Facility Test, the Oral Language Evaluation, and the Pictorial Test of Bilingualism and Lan-

Note that these three grammatical morphemes are all typically manifested by the sound of s and are thereby auditorily equally salient to the perceptions of the language learner. This case is one example of the third type of evidence mentioned above for the existence of a natural sequence in the acquisition process.

guage Dominance require children to describe or tell a story about pictures. A few tests, such as the Language Dominance Criteria, educe conversation in the form of answers to personal questions. Discrete point tests, on the other hand, consist of items designed to test a number of specific structures or rules--discrete points in the language system. Generally, each item tests one such point, independently of other items which test different points. In such a test, all the items taken together are considered to be a representative sampling of the structure of the language, or at least of some part of its structure--grammar, phonology, etc.

Integrative Tests. Because integrative tests have the virtue of eliciting natural discourse, the language produced in such tests will reflect the acquired knowledge of English structure; but, of the integrative tests reviewed, none yields any useful information regarding the degree of control of grammatical structure or developmental level of English acquisition.

A general problem with integrative testing is that a sample of natural discourse will provide information only about the grammatical structures which happen to occur in the sample. It will provide a basis on which to assess the control of those structures, but it will not provide information about the control of any structures which do not occur. We will learn something about control of irregular past tense from the student who, for example, utters either *He took a nap* or *He taked a nap*, but we will have no basis for making such a judgment if, in the discourse, no irregular verbs occur in past tense. This fact limits the evaluative and diagnostic potential of integrative tests.

A second and greater problem with the integrative tests we reviewed lies in the nature of their scoring procedures. The types of evaluations which are made of elicited discourse are either so gross as to be unrevealing, so subjective as to limit their value, or based on misguided notions about language. One example of gross and unrevealing scoring procedures is provided by the Language Dominance Criteria, where the scorer simply indicates the presence or absence of grammatical errors in the discourse. Similarly, on the Chicago Short Test of Linguistic Skills, the scorer indicates whether there are major or minor grammatical errors, with no more than a few ungeneralizable examples to guide judgment. Examples of extremely subjective evaluations can be found in the Language Facility Test, in which the examiner judges the organization and descriptive adequacy of the discourse; in the Oral Language Evaluation, in which the discourse is judged in terms of concrete versus abstract content; or in the Language Assessment Scales, in which language production is judged to be incoherent, labored, near perfect, or perfect. (The last category is an ideal which not even native speakers would characteristically achieve.) Other tests, such as the Basic Inventory of Natural Language, Spanish/English Language Performance Screening test, or Test of Grammatically Correct Spanish/English, involve scoring of sentence length or complexity. Whatever value may lie in some of these scoring procedures, it is clear that none yields diagnostic information about students' control of English structure, and none indicates developmental level of English acquisition.

Discrete Point Tests. One of the advantages of discrete point testing (which enables it to overcome or avoid the problems associated with integrative testing) is that it can, in principle, assure controlled representation of structures or rules which, for one reason or another, are considered important. However, one of the main disadvantages of discrete point tests is that it is difficult to extract natural language responses while maintaining close control over the grammatical structures to be produced.

There are two groups of discrete point grammar tests whose tasks do not cause students to produce any spontaneous natural language, and which will therefore fail to yield information regarding students' developmental level of English acquisition. The first group includes tests of aural comprehension, such as the Comprehension of

Oral Language Test, or sections of the Language Assessment Scales, the MAT-SEA-CAL, the Language Assessment Battery, or the Milwaukee English as a Second Language Achievement Tests. As previously discussed, it is very difficult to assess control of grammatical structure by trying to evaluate passive comprehension. Furthermore, aural comprehension does not necessarily reflect students' ability to speak the language.

The second group of tests which does not involve elicitation of spontaneous natural language includes tests with sentence repetition tasks, such as the Northwestern Syntax Screening Test or the Del Rio Language Screening Test. It should be mentioned that both the Northwestern and the Del Rio tests are designed to identify children with delayed or deviant language development. The manual of the Northwestern test (to its credit) states that "a bilingual child...should never be judged as 'language delayed' or 'language deficient' by his scores on the NSST." In contrast, the manual of the Del Rio test says that "though not originally intended for use in determining a child's degree of bilingualism, the test may be used for this purpose--a useful and needed additional benefit." Neither test, in fact, can usefully indicate a normal bilingual child's developmental level in English acquisition.

Items in most of the discrete point grammar tests reviewed evoke a structured response of some type. Some, such as the Michigan Oral Language Productive Test, the MAT-SEA-CAL, and the Language Assessment Battery, offer a statement or question "cue," followed by the beginning of a response sentence which students are expected to complete. This task seems to decrease the likelihood that responses will demonstrate relatively unmonitored, spontaneous natural language production. Other tests, such as the Moreno Oral English Proficiency Placement Test, the Ilyin Oral Interview, the English as a Second Language Achievement Tests, the Caso Test for Non-English Speaking Students, the Bilingual Syntax Measure, the Bahia Oral Language Test, and the Oral Language Dominance Measure, simply ask questions to which students respond. Certain features in the design of these tests prevent many of the test items from eliciting natural language.

Beyond the technical design problems, the degree to which structured response items will succeed in obtaining natural language depends not only on the content of questions asked, but on such difficult-to-control factors as the testing situation or the skill of the examiner in creating a relaxed situation. The less test items resemble ordinary conversation, and the more formal the testing situation, the more likely it is that student responses will reflect a high degree of self-monitoring. One consequence of this formal, self-monitored situation might be to produce reticence or anxiety on the part of the student and consequently a poor picture of language ability. Another outcome might be that, in many cases, the test could yield evidence of learned grammar rather than the internalized, acquired language knowledge which enables the student to use English for communication and learning. Among discrete point tests, the Bilingual Syntax Measure, the Bahia Oral Language Test, and the Oral Language Dominance Measure appear to have a high potential, properly administered, for eliciting natural language.

As we have tried to show, neither discrete point nor integrative testing has an overriding advantage over the other for purposes of assessing students' developmental level of English acquisition. Integrative testing has the virtue of eliciting natural discourse, but in order to be diagnostic of language development, its scoring procedures would require a type of analysis of the discourse which no test currently incorporates. Discrete point testing can ensure coverage of structures needed for assessment, but often fails to educe the natural language which would reflect students' placement in the developmental process. Both approaches, in principle, hold potential for improvement in further test development.

Correspondence between Test Content and Acquisition Order

Some structured-response, discrete point tests of grammar do seem to have potential for eliciting natural language. In evaluating these tests further, we need to ask what correspondence exists between linguistic structures tested, and which ones indicate the natural acquisition order of English (to the extent documented by research).

This is a crucial question to ask of any test claiming to be representative of English grammatical structure. The tests of this nature which we reviewed are all based--implicitly or explicitly--on some hierarchical list of supposedly representative grammatical structures or, in some cases, concepts. Such hierarchical lists will claim to reflect one of two things: first, they may claim to represent the way the English language really is, or second, they may correspond to some artificial classification of English structure, such as a specific instructional curriculum. In evaluating a discrete point grammar test, we have to consider what motivated the selection of structure on the test (the content of its items) in order to know what a score on that test reveals about students' language ability. If the hierarchy of structures on a test reflects the natural acquisition order of English structure, then performance on the test will tell us something about development in English. If the test content does not reflect that order, we have to ask what it does tell us about the student, and of what value that information is. The following discussion will illustrate these points.

Items on the Moreno Oral English Proficiency Placement Test correspond in content and sequence to a particular set of curriculum materials (available from the same company which markets the test), but do not correspond to any acquisition order revealed by language acquisition research. For example, research indicates that the plural copula is a relatively early acquisition, while possessives are acquired later. But on the Moreno test, items to test possessives occur early, while items inducing the plural copula occur late, and the scoring procedure might prevent the child who misses the possessives from ever getting to the copula items. Similarly, the Moreno test sequences irregular past tenses before progressives, the reverse of the natural order of difficulty. Students' scores on this test, then, will indicate place in the particular curriculum on which the test is based, but will not indicate much about actual English proficiency. In the case of curriculum-based tests like this, divergence from the natural order of acquisition is a problem properly located in the curriculum rather than in the test.

For several tests (or the grammar sections thereof), such as the Ilyin Oral Interview, the Caso Test for Non-English Speaking Students, the Milwaukee English as a Second Language Achievement Tests, or Language Assessment Battery, it is not clear why the item content--or the weighting of different structures--is what it is. On the Milwaukee ESL Tests, for example, over half the questions extract responses of the form *noun + be + predicate*. Scores on this test will be determined in large part by students' control of this particular type of structure, while control of other aspects of grammar will not be indicated. On the Language Assessment Battery, there are items which elicit structures whose role in the system of English is relatively idiosyncratic, such as the prepositional phrase *of + noun* in the stimulus sentence *It is a telephone booth full* _____. Such an item, which does not indicate control of any general rule of the language, contributes little to an assessment of proficiency.

Items on the MAT-SEA-CAL test also bear no apparent relationship to the acquisition order of English, and reflect no formally motivated analysis of the structure of English. Items are designed to test skills in eight concept areas. Some of these concepts represent clusters of grammatical forms (e.g., temporality, including verb tense and aspect), while others seem to be semantic or cognitive categories (e.g., classification or identification), which are not grammatically motivated. Scores

which are categorized in terms of these concepts will indicate very little about students' control of English structure.

Item content on the Michigan Oral Language Productive Test appears designed to provide coverage of known features of nonstandard varieties of English, particularly Black Vernacular English). For example, one category of items tests for double negative; if the test were truly intended to reveal the English proficiency of non-native speakers, there would be no reason to test for this kind of linguistic feature, because there is no developmental stage in the acquisition of English which is characterized by the presence of double negatives. This is, rather, a feature of the nonstandard dialect of English spoken by many persons with native proficiency. This assessment of the Michigan test is borne out by the scoring procedures, which give credit to specified standard English responses, lumping together nonstandard and other responses. This test, in other words, equates those whose responses reflect native proficiency in a nonstandard variety of English with non-native speakers of English whose responses indicate they are in the process of acquiring English as a second language. Yet, the knowledge of English is fundamentally different--qualitatively and quantitatively--in these two cases. The Michigan test will reveal little about the developmental stage of students acquiring English as a second language.

The content of the Bilingual Syntax Measure (BSM), the Bahia Oral Language Test (BOLT), and the El Paso Oral Language Dominance Measure represents a number of English structures for which some order of difficulty has been established by research. To this extent, then, there is a good rationale for the claim that these tests assess English proficiency. In the BSM (remarks here refer to the "hierarchical scoring method") and BOLT, however, scoring procedures introduce some possibility that the tests might fall short of that goal. In these tests (as well as in the El Paso test), each item is designed to elicit an utterance which will contain a particular type of structure--the *target structure* for that item. Responses are acceptable if they are grammatical and appropriate in the context of the given cue question. In the BSM and BOLT, different subsets of items are taken to represent different levels of proficiency. Thus, students' proficiency levels according to the test depend not on a numerical total of acceptable responses, but on performance on the specific subsets of items. The rationale for this is that the target structures for the items are located at different points in the acquisition hierarchy, so that students' performance on certain items will indicate placement in the developmental sequence.

This procedure rests on two assumptions, neither of which seems entirely safe to make: (a) each item will, in fact, elicit its target structure (rather than an acceptable response which does not contain the target structure) and (b) if responses to an item are unacceptable, this will be because the target structure, rather than some other part of the response, is ill formed. If either of these assumptions fails to hold--which must sometimes be the case--the scoring procedure will be robbed of the basis for relating scores on specific items to proficiency levels.

A further problem with these tests is that they place a very heavy reliance on the details of the acquisition hierarchy which they are designed to represent. Some investigators would argue that it is premature to trust the research results on acquisition sequence beyond a certain level of detail. Furthermore, these tests make the implicit claim that the acquisition hierarchy that has been uncovered is generalizable to every child learning English. They do not allow for the possibility that there may be individual variation within the general acquisition sequence.

The El Paso test yields a score which represents the numerical total of acceptable responses. In this, it avoids problems discussed above, but perhaps at some cost of diagnostic potential. Translation of scores on the El Paso test into proficiency levels is based on statistical analysis of scores from a local field test population, and this analysis would have to be repeated for other local populations before the test could be used to provide proficiency level data in other locations.

SUGGESTED READINGS

Anderson, J. "Order of Difficulty in Adult Second Language Acquisition." In Second Language Acquisition Research. New York: Academic Press, 1978.

Burt, M. and H. Dulay. "Some Guidelines for the Assessment of Oral Language Proficiency and Dominance." TESOL Quarterly 12:2, 1978.

Burt, M., H. Dulay, and E. Hernández-Chávez. Bilingual Syntax Measure, Technical Handbook. Harcourt, Brace, Jovanovich, Inc., 1976.

Cazden, C., H. Cancino, E. Rosansky, and J. Schumann. Second Language Acquisition Sequences in Children, Adolescents and Adults. Final Report to U.S. Dept. of Health, Education and Welfare National Institute of Education, Project No. 730744, 1975. ED 121 115

Chomsky, C. The Acquisition of Syntax in Children from 5 to 10. Cambridge: The M.I.T. Press, 1969.

d'Anglejan, A. and G.R. Tucker. "The Acquisition of Complex English Structures by Adult Learners." Language Learning 25:2, 1975.

de Villiers, J. and P. de Villiers. Language Acquisition. Cambridge: Harvard University Press, 1978.

Dulay, H. and M. Burt. "Natural Sequences in Child Second Language Acquisition." Language Learning 24:1, 1974.

Dulay, H. and M. Burt. "A New Perspective on the Creative Construction Process in Child Second Language Acquisition." Language Learning 24:2, 1974.

Dulay, H. and M. Burt. "Creative Construction in Second Language Learning and Teaching." In Papers in Second Language Acquisition, edited by H. Brown. Language Learning, special issue no. 4, 1976.

Dulay, H. and M. Burt. "Some Remarks on Creativity in Language Acquisition." In Second Language Acquisition Research: Issues and Implications, edited by W. Ritchie. New York: Academic Press, 1978.

Edmonds, M. "New Directions in Theories of Language Acquisition." Harvard Educational Review 46:2, 1976.

Ervin-Tripp, S. Language Acquisition and Communicative Choice. Stanford: Stanford University Press, 1973.

Fillmore, L. "Individual Differences in Second Language Acquisition." Paper presented to Asilomar Conference on Individual Differences in Language Ability and Language Behavior, Monterey, CA, Nov. 1975.

Fathman, A. "The Relationship between Age and Second Language Productive Ability." Language Learning 25:2, 1975.

Hakuta, K. and H. Cancino. "Trends in Second-Language-Acquisition Research." Harvard Education Review 47:3, 1977.

Hatch, E. "Second Language Learning." In Bilingual Education: Current Perspectives/ Linguistics, Vol. 2. Arlington: Center for Applied Linguistics, 1977. ED 142 073

_____. "Discourse Analysis, Speech Acts, and Second Language Acquisition." In Second Language Acquisition Research: Issues and Implications, edited by W. Ritchie. New York: Academic Press, 1978.

_____, ed. Second Language Acquisition. Rowley, MA: Newbury House, 1978.

Hyltenstam, K. "Implicational Patterns in Interlanguage Syntax Variation." Work in Progress No. 10, University of Edinburgh, 1977.

Ioup, G. and A. Kruse. "Interference versus Structural Complexity in Second Language Acquisition: Language Universals as a Basis for Natural Sequencing." In On TESOL '77: Teaching and Learning English as a Second Language: Trends in Research and Practice, edited by H. Brown, C. Yorio, and R. Crymes. Washington, DC: TESOL, 1977.

Krashen, S. "The Monitor Model for Second-Language Acquisition." In Second-Language Acquisition and Foreign Language Teaching, edited by R. Gingràs. Arlington: Center for Applied Linguistics, 1978.

_____, V. Sferlazza, L. Feldman, and A. Fathman. "Adult Performance on the SLOPE Test: More Evidence for a Natural Sequence in Adult Second Language Acquisition." Language Learning 26:1, 1976.

Krashen, S., N. Houck, P. Giunchi, S. Bode, R. Birnbaum, and J. Strei. "Difficulty Order for Grammatical Morphemes for Adult Second Language Performers Using Free Speech." TESOL Quarterly 11:3, 1977.

Larsen-Freeman, D. "The Acquisition of Grammatical Morphemes by Adult ESL Students." TESOL Quarterly 9:4, 1975.

_____. "Evidence of the Need for a Second Language Acquisition Index of Development." In Second Language Acquisition Research: Issues and Implications, edited by W. Ritchie. New York: Academic Press, 1978.

_____. "An Explanation for the Morpheme Accuracy Order of Learners of English as a Second Language." In Second Language Acquisition, edited by E. Hatch. Rowley, MA: Newbury House, 1978.

McLaughlin, B. Second-Language Acquisition in Childhood. New York: John Wiley and Sons, 1978.

Richards, J. "Error Analysis and Second Language Strategies." Language Sciences 17, 1971.

Ritchie, W., ed. Second Language Acquisition Research: Issues and Implications. New York: Academic Press, 1978.

Rosansky, Ellen. "Methods and Morphemes in Second Language Acquisition Research." Language Learning 26:2, 1976.

Sheldon, A. Assumptions, Methods, and Goals in Language Acquisition Research. Bloomington: Indiana University Linguistics Club, 1978.

Wagner-Gough, J. and E. Hatch. "The Importance of Input Data in Second Language Acquisition Studies." Language Learning 25:2, 1975.

[Documents identified by an ED number may be read on microfiche at an ERIC library collection or ordered from the ERIC Document Reproduction Service, P.O. Box 190, Arlington, VA 22210.]

Part Two

CATALOGUE OF TESTS

Part Two is an alphabetical catalogue of the 38 tests reviewed. It provides certain general information about each test and adds some evaluative comments which may not have been covered in Part One. It also serves as an index to Part One, citing where a particular test is discussed or where examples drawn from particular tests are used to illustrate issues in language proficiency. In cases where a test was not mentioned in connection with a particular topic, the index references the discussions that are useful for evaluating the test. Readers who are considering using any of these tests should consult these sections.

Additionally, the catalogue contains information about what the test is intended to measure, either as stated in the test's technical materials or, if not stated, as we determined it through analysis of the test. A statement of what tasks are actually required of people taking the test (e.g., answering questions, pointing at pictures, performing commands, etc.) is also provided.* We believe that it is essential to know these details in order to get an idea of what the test is like, and what it would be like to administer the test in an educational setting.

Information about the existence or outcome of experimental studies of reliability and validity for each test has not been included. We have also not included any judgments about the acceptability or unacceptability of any test, either in general or with respect to particular topics of interest, such as cultural fairness, layout and design, age/grade appropriateness, ease of administration, or cost economy. These issues we consider to be beyond the scope of the present review, and we believe that categorical judgments of acceptability in the absence of any explanation cannot be regarded as very useful or trustworthy information. Technical and statistical data, results of experimental field testing, and judgments concerning various aspects of overall acceptability are available in a number of previously published test catalogues and reviews. Some of those publications** now available are:

Backer, Thomas E. A Directory of Information on Tests (TM Report 62). ERIC Clearinghouse on Tests, Measurement and Evaluation, 1977. ED 152 802

Bye, Thomas T. Tests that Measure Language Ability. Berkeley, CA: BABEL/LAU Center, Berkeley Unified School District, 1977.

The publication Tests that Measure Language Ability contains information about what each test measures and what tasks are required. We acknowledge their lead in our decision to provide this information in our own review.

***These catalogues vary widely in terms of the type and amount of information they offer; some will be more informative and useful than others to school personnel faced with the task of selecting among tests. The list is presented here for informative purposes only, and inclusion of any publication on this list does not necessarily represent an endorsement or recommendation of that publication by the present authors.*

California Association for Bilingual Education. A Qualitative Review of Instruments Used in Bilingual Education Programs in California. San Jose, CA: Test Evaluation Conference, April 1976.

Grommon, Alfred H., ed. Reviews of Selected Published Tests in English. Urbana, IL: National Council of Teachers of English, 1976. ED 120 732

Locks, Nancy, Barbara A. Pletcher, and Dorothy F. Reynolds. Language Assessment Instruments for Limited-English-Speaking Students: A Needs Analysis. Washington, DC: National Institute of Education, 1978. ED 163 062

Oller, John W. and Kyle Perkins. Language in Education: Testing the Tests. Rowley, MA: Newbury House, 1978.

Pletcher, Barbara and others. A Guide to Assessment Instruments for Limited-English-Speaking Students. New York: Santillana Publishing Company, 1978.

Savard, Jean-Guy. Analytical Bibliography of Language Tests. Québec: Les Presses de l'Université Laval, 1977. ED 032 540

Silverman, Robert J., Joslyn K. Noa, and Randall H. Russell. Oral Language Tests for Bilingual Students: An Evaluation of Language Dominance and Proficiency Instruments. Portland, OR: Northwest Regional Educational Laboratory, 1976. ED 129 099

* * *

BAHIA ORAL LANGUAGE TEST

AUTHORS:	Sam Cohen, Roberto Cruz, Raul Bravo
AVAILABLE FROM:	Bay Area Hispano Institute for Advancement (BAHIA, Inc), PO Box 9337, North Berkeley, CA 94709
DATE: 1976	AGE/GRADE LEVEL: grades 7-12
TEST INTENDS TO MEASURE:	Oral proficiency by testing control of specific grammatical structures
TASKS INVOLVED:	Students answer examiner's questions about pictures
INDEX TO PART ONE:	Ellipsis and Complete Sentences; Identical Items in Two Languages; Discrete Point vs. Integrative Testing; Correspondence between Test Content and Acquisition Order

BASIC INVENTORY OF NATURAL LANGUAGE

AUTHORS:	Charles H. Herbert
AVAILABLE FROM:	CHECpoint Systems, 1558 N. Waterman Ave, Suite C, San Bernardino, CA 92404
DATE: 1977	AGE/GRADE LEVEL: preschool-adult
TEST INTENDS TO MEASURE:	Fluency (total number of words used); level of complexity (command of linguistic structures, including sentences, modifiers, phrases, and clauses)
TASKS INVOLVED:	Students describe or tell a story about a picture, producing an extended sample of connected speech (the sample should be produced in a conversational context or interchange with other students)

INDEX TO PART ONE: Notions of Linguistic Complexity; Counting Linguistic Elements; Ellipsis and Complete Sentences; Identical Items in Two Languages; Discrete Point vs. Integrative Testing

Re Linguistic Complexity: A "Grammatical Index" serves as the basis for one of the scores on this test, and makes certain commitments about the relative complexity of various linguistic forms. The relative weights assigned do not seem to be well motivated, either from the point of view of linguistics or second language acquisition.

BILINGUAL SYNTAX MEASURE

AUTHORS: Marina K. Burt, Heidi C. Dulay, Eduardo Hernández-Chávez

AVAILABLE FROM: Test Department, Harcourt, Brace, Jovanovich, Inc, 757 Third Ave, New York, NY 10017

DATE: 1975 AGE/GRADE LEVEL: grades K-2

TEST INTENDS TO MEASURE: Control of specific grammatical structures related to the acquisition sequence in English grammar

TASKS INVOLVED: Students answer examiner's questions about pictures

INDEX TO PART ONE: Ellipsis and Complete Sentences; Identical Items in Two Languages; Discrete Point vs. Integrative Testing; Correspondence between Test Content and Acquisition Order

BROOKS COUNTY LANGUAGE USAGE INVENTORY

AUTHORS:

AVAILABLE FROM: Brooks County Independent School District, Falfurrias, TX 78355

DATE: None AGE/GRADE LEVEL: Not indicated

TEST INTENDS TO MEASURE: General comprehension; language preference (in particular social situations); vocabulary

TASKS INVOLVED: Students answer questions about themselves, the examiner's actions, and the language used or preferred in various social situations; students point to pictures in response to the examiner's instructions

INDEX TO PART ONE: Self-Reported Data; Understanding the Test Task; Identical Items in Two Languages

Re Understanding the Test Task: In this section, we criticized some tests for implicitly assuming that all students will have an explicit awareness that a certain way of speaking is called "English," while a different way is called "Spanish." The Inventory recognizes this and recommends some preliminary discussion to ensure that students understand when the examiner refers to English and/or Spanish.

CASO TEST FOR NON-ENGLISH SPEAKING STUDENTS

AUTHORS: Adolph Caso

AVAILABLE FROM: Adolph Caso, Waltham Public Schools, Waltham, MA 02154

DATE: 1975 AGE/GRADE LEVEL: grades 3-8

TEST INTENDS TO MEASURE: Phonetics and spelling; vocabulary; reading and listening comprehension

TASKS INVOLVED: Writing letters and words from dictation; identifying synonyms, antonyms, and related words; multiple-choice questions about a reading passage; multiple-choice questions about a dictated passage

INDEX TO PART ONE: Tasks that Require Literacy Skills; The Value of Vocabulary Tests; Measures of Linguistic Creativity; Ellipsis and Complete Sentences; Metalinguistic Instructions; Discrete Point vs. Integrative Testing; Correspondence between Test Content and Acquisition Order

Re Literacy Skills: Two parts of the test require literacy skills in addition to oral English proficiency. Part I (phonetics) requires students to write letters and spell words from dictation; Part II (comprehension) requires students to read words and identify synonyms and antonyms from written lists.

Re Metalinguistic Instructions: Spelling ability and the recognition of synonyms and antonyms are skills that are based on an explicit analysis of language rather than skills that properly constitute language proficiency. These are metalinguistic tasks.

COMPREHENSION OF ORAL LANGUAGE

AUTHORS: Not indicated

AVAILABLE FROM: Guidance Testing Associates, St. Mary's University, 2700 Cincinnati Ave, San Antonio, TX 78284

DATE: None AGE/GRADE LEVEL: Not indicated

TEST INTENDS TO MEASURE: Comprehension of words and certain grammatical structures

TASKS INVOLVED: Students identify pictures corresponding to words and expressions presented orally by the examiner

INDEX TO PART ONE: Tasks that Reduce to a Vocabulary Test; Passive Comprehension; The Value of Vocabulary Tests; Discrete Point vs. Integrative Testing

Re Tasks that Reduce to a Vocabulary Test: Students identify pictures corresponding to sentences spoken by the examiner. On many of these items, students can identify the correct picture by understanding only a word or two, and need not understand the whole sentence. It is not certain that such items test sentence comprehension; rather, they may tap only vocabulary knowledge.

Re Discrete Point vs. Integrative Testing: See especially the discussion in this section regarding the inadequacy of aural comprehension tests for indicating students' developmental level of English acquisition.

CRANE ORAL DOMINANCE TEST

AUTHORS: Barbara J. Crane

AVAILABLE FROM: Crane Publishing Co, 1301 Hamilton Ave, PO Box 3713, Trenton, NJ 08629

DATE: 1976 (reviewed) AGE/GRADE LEVEL: 4-8 years
1978 revision available

TEST INTENDS TO MEASURE: Dominant internal language; memory for English and Spanish vocabulary items

TASKS INVOLVED: Examiner orally presents a list of eight words to students, four in English, four in Spanish; students recall as many words as possible in either language

INDEX TO PART ONE: Tasks that Reduce to a Vocabulary Test; Mimicry; Understanding the Test Task; Metalinguistic Instructions

Re Mimicry: The memory task imposed by this test requires students to repeat words spoken by the examiner. Students usually are able to remember more words in a given language if they understand the words; however, the ability to mimic words, without necessarily understanding them, cannot be ruled out as a factor affecting performance on this test.

DEL RIO LANGUAGE SCREENING TEST

AUTHORS: Allen S. Toronto, D. Leverman, Cornelia Hanna, Peggy Rosenzweig, Antoneta Maldonado

AVAILABLE FROM: Not indicated

DATE: 1975 AGE/GRADE LEVEL: 3-6 years

TEST INTENDS TO MEASURE: Receptive vocabulary; sentence repetition or memory; comprehension of oral commands; story comprehension

TASKS INVOLVED: Students identify pictures corresponding to words presented orally; the examiner presents a series of sentences of approximately equal complexity, but increasing length, and students recall and repeat the sentences; students perform oral commands presented by the examiner; the examiner reads orally a series of five stories increasing in length and complexity, and students answer questions presented orally after each story

INDEX TO PART ONE: Notions of Linguistic Complexity; Mimicry; The Value of Vocabulary Tests; Ellipsis and Complete Sentences; Discrete Point vs. Integrative Testing

Re Ellipsis and Complete Sentences: In scoring the "Story Comprehension" subtest, normal, elliptical answers to examiner's questions are counted correct. This test avoids the errors pointed out in this section.

ENGLISH AS A SECOND LANGUAGE ACHIEVEMENT TESTS--LEVEL I

AUTHORS: Helen McGuire, Susan Rao

AVAILABLE FROM: Milwaukee Public Schools, Department of Education Research and Program Assessment, PO Drawer 10-K, Milwaukee, WI 53201

DATE: 1978 AGE/GRADE LEVEL: grades K-12 (listening and speaking)
grades 6-12 (reading and writing)

TEST INTENDS TO MEASURE: Listening comprehension, correct pronunciation and intonation; reading (not reviewed); writing (not reviewed)

TASKS INVOLVED: Students identify pictures corresponding to sentences and "stories"

(sequences of several sentences) presented orally; students repeat words and sentences presented orally by the examiner; students answer personal and general questions

INDEX TO PART ONE: Tasks that Reduce to a Vocabulary Test; Mimicry; Passive Comprehension; The Value of Phonology or Pronunciation Tests; Counting Linguistic Elements; Ellipsis and Complete Sentences; Discrete Point vs. Integrative Testing; Correspondence between Test Content and Acquisition Order

Re Passive Comprehension: With listening comprehension tasks in which students do not actually produce speech, it is often difficult to claim that certain test items really test students' understanding of a particular grammatical structure; understanding one or a few words is often sufficient to get items correct.

Re The Value of Phonology or Pronunciation Tests: The "Speaking" subtest requires repetition of words or sentences spoken by the examiner, and is scored for correct pronunciation. This test is subject to the criticisms we presented in this section.

Re Counting Linguistic Elements: The sentence repetition task on the "Speaking" subtest is scored in part by counting the number of syllables students produce and matching it with the number of syllables in the test sentence. It is unclear what can be inferred about language proficiency from students' ability to match the number of syllables in a spoken sentence.

FLEXIBILITY TEST TO MEASURE DOMINANCE IN SPANISH-ENGLISH BILINGUALS

AUTHORS: Gary D. Keller

AVAILABLE FROM: Educational Testing Service, ETS Test Collection, Princeton, NJ 08540 (microfiche)

DATE: 1974 AGE/GRADE LEVEL: grades 5 and 6

TEST INTENDS TO MEASURE: Spanish-English dominance

TASKS INVOLVED: Students are given a nonsense sequence of eight letters and asked to construct as many words as possible from those letters, first in Spanish, then in English; this is repeated five times with different letters

INDEX TO PART ONE: Tests that Require Literacy Skills; Tasks that Reduce to a Vocabulary Test; Understanding the Test Task; Metalinguistic Instructions

Re Literacy Skills: The format of this test--spelling words from scrambled letters--is heavily dependent on literacy skills in addition to linguistic proficiency.

Re Tasks that Reduce to a Vocabulary Test: Beyond reading and spelling ability, the one linguistic skill tested is vocabulary knowledge; compare other kinds of test tasks that also reduce to vocabulary tests as discussed in this section.

Re Understanding the Test Task: The scrambled letter task is a kind of language game only indirectly related to ordinary language use. If bilingual students score low on both English and Spanish, one reasonable explanation would be that they have not understood the rules of the game being played.

HARRIS ARTICULATION TEST

AUTHORS: Gail Harris

AVAILABLE FROM: Educational Testing Service, ETS Test Collection, Princeton, NJ 08540 (microfiche)

DATE: 1974 AGE/GRADE LEVEL: grades preK-2

TEST INTENDS TO MEASURE: Pronunciation

TASKS INVOLVED: Students name objects represented in pictures, producing particular sounds in initial, medial, and final positions

INDEX TO PART ONE: Tasks that Reduce to a Vocabulary Test; Mimicry; The Value of Phonology or Pronunciation Tests; Metalinguistic Instructions

Re Tasks that Reduce to a Vocabulary Test: Although this is graded as a pronunciation test, there is also a vocabulary component to it. Students must name pictures of objects; if the English names of some of the objects are not known, the test will be unable to assess whether certain sounds in certain environments can be pronounced. On the other hand, the method of naming pictures does avoid the problem of mimicry, which pronunciation tests often encounter.

HAYWARD LANGUAGE DOMINANCE INDICATOR

AUTHORS: Office of Bilingual Multicultural Education, Educational Services Division, Hayward Unified School District

AVAILABLE FROM: Hayward Unified School District, PO Box 5000, Hayward, CA 94540

DATE: None AGE/GRADE LEVEL: grades 1-6

TEST INTENDS TO MEASURE: Vocabulary; language use in particular social situations

TASKS INVOLVED: The examiner mentions a location, and students think of objects found in that location; students answer questions about what language they use with various family members and in certain social situations

INDEX TO PART ONE: Self-Reported Data; The Value of Vocabulary Tests; Counting Linguistic Elements; Identical Items in Two Languages

Re Identical Items in Two Languages: On the vocabulary section of this test, students produce a list of objects found in various locations--in the kitchen, in the street--first in Spanish, then in English. As argued here in Part One, it is unnatural from the point of view of ordinary language use to ask someone to repeat the same information twice.

ILYIN ORAL INTERVIEW

AUTHORS: Donna Ilyin

AVAILABLE FROM: Newbury House Publishers, 54 Warehouse Rd, Rowley, MA 01969

DATE: None AGE/GRADE LEVEL: 12+ years

TEST INTENDS TO MEASURE: Oral English production

TASKS INVOLVED: Answering (and asking) questions about a series of pictures; use of specific words or structures is sometimes required

INDEX TO PART ONE: Measures of Linguistic Creativity; Understanding the Test Task; Ellipsis and Complete Sentences; Metalinguistic Instructions; Discrete Point vs. Integrative Testing; Correspondence between Test Content and Acquisition Order

JAMES LANGUAGE DOMINANCE TEST

AUTHORS: Peter James

AVAILABLE FROM: Learning Concepts, 2501 N. Lamar, Austin, TX 78705

DATE: 1975 AGE/GRADE LEVEL: grades K and 1

TEST INTENDS TO MEASURE: Vocabulary comprehension; vocabulary production

TASKS INVOLVED: Identifying pictures corresponding to words presented orally by the examiner; naming objects or activities represented in pictures pointed to by the examiner

INDEX TO PART ONE: The Value of Vocabulary Tests; Metalinguistic Instructions; Identical Items in Two Languages

Re Identical Items in Two Languages: This test requires students to identify pictures and name objects in English and Spanish. The same pictures are used for both languages, but not in the same order. The criticisms regarding duplicate items in two languages apply to this test.

LANGUAGE ASSESSMENT BATTERY

AUTHORS: Office of Educational Evaluation, New York City Board of Education

AVAILABLE FROM: Houghton-Mifflin Co, Test Department, 777 California Ave, Palo Alto, CA 94304

DATE: None AGE/GRADE LEVEL: grades K-2 (Level 1); grades 3-6 (Level 2); grades 7-12 (Level 3)

TEST INTENDS TO MEASURE: Speaking; listening; reading (not reviewed); writing (not reviewed)

TASKS INVOLVED: Answering personal questions, naming body parts touched by examiner, naming objects in pictures; identifying body parts named by the examiner, identifying pictures of objects named by the examiner (Level 1); answering the examiner's question, or completing the examiner's probe sentences about pictures (Levels 2 and 3); identifying pictures of objects named by the examiner (Levels 2 and 3); among three pairs of words, identifying the one pair pronounced exactly the same (Levels 2 and 3)

INDEX TO PART ONE: Tasks that Reduce to a Vocabulary Test; Passive Comprehension; The Value of Vocabulary Tests; The Value of Phonology or Pronunciation Tests; Ellipsis and Complete Sentences; Discrete Point vs. Integrative Testing; Correspondence between Test Content and Acquisition Order

Re Passive Comprehension: On the Level 1 test, all the items requiring students to speak require only single-word answers. All the items that attempt to assess students' ability to understand

grammatical structure are listening items in which students identify pictures corresponding to what the examiner says. Thus, the test depends on evaluating passive listening comprehension in assessing students' control of grammatical structure. Levels 2 and 3 include speaking tasks which require production of extended expressions and grammatical structures, and so are not subject to the same criticisms covered in this section.

LANGUAGE ASSESSMENT SCALES

AUTHORS: Sharon E. Duncan, Edward A. DeAvila

AVAILABLE FROM: DeAvila, Duncan and Associates, PO Box 770, Larkspur, CA 94939

DATE: 1977 (Level 1) 1978 (Level 2) AGE/GRADE LEVEL: grades K-5 (Level 1); grades 6+ (Level 2)

TEST INTENDS TO MEASURE: Perception and pronunciation of English sounds; vocabulary; sentence comprehension; oral production (paraphrase)

TASKS INVOLVED: Students identify pairs of words presented orally as being the same or different; students name objects represented in pictures; students repeat words and sentences presented orally; students identify pictures corresponding to sentences presented orally; students repeat and paraphrase an orally presented story describing a series of pictures

INDEX TO PART ONE: Tasks that Reduce to a Vocabulary Test; Mimicry; Passive Comprehension; The Value of Vocabulary Tests; The Value of Phonology or Pronunciation Tests; Discrete Point vs. Integrative Testing

LANGUAGE DOMINANCE CRITERIA

AUTHORS: Fernando Canedo, David Gustafson, Américo López-Rodríguez

AVAILABLE FROM: Bilingual-Bicultural Education Program, California State University, Fullerton, CA 92634

DATE: None AGE/GRADE LEVEL: grades K and 1

TEST INTENDS TO MEASURE: Language spoken by students and their family members; speaking ability, fluency and grammar

TASKS INVOLVED: Students answer questions about what language they speak with particular family members and about what language family members speak most; students answer personal questions, and, if short answers are given, the examiner asks further questions about what students have said

INDEX TO PART ONE: Self-Reported Data; Identical Items in Two Languages; Discrete Point vs. Integrative Testing

Re Identical Items in Two Languages: The test includes a sociolinguistic questionnaire, which students answer in Spanish if they are able to do so, and a performance section in which students answer discussion questions in Spanish. They later answer similar questions in English. Since the questions are not simply translations, the test avoids the unnaturalness of making students answer the same question twice.

Re Discrete Point vs. Integrative Testing: See especially the discussion of this test here.

LANGUAGE DOMINANCE SURVEY

AUTHORS:	San Bernardino Bilingual Office
AVAILABLE FROM:	San Bernardino City Unified School District, 799 F St, San Bernardino, CA 92410
DATE: 1974-75	AGE/GRADE LEVEL: grades K-1, 2-4, 5-7, 8-12 (separate forms)
TEST INTENDS TO MEASURE:	Listening and speaking (K-1); for all other forms, listening comprehension; speaking proficiency; reading comprehension (not reviewed); writing proficiency (not reviewed)
TASKS INVOLVED:	Students identify pictures in response to sentences presented orally and answer questions about pictures (K-1); students answer questions and perform commands presented orally as well as answer personal questions and produce verbal responses to a command (2-4); students perform commands presented orally, answer personal questions, and produce verbal responses to a command (5-7, 8-12)
INDEX TO PART ONE:	Self-Reported Data; Identical Items in Two Languages; Discrete Point vs. Integrative Testing Re Identical Items in Two Languages: The listening and speaking sections of this test are presented in English and Spanish. The questions and instructions are different in each language, so the criticisms developed about duplicate items are avoided.

LANGUAGE FACILITY TEST

AUTHORS:	John T. Dailey
AVAILABLE FROM:	The Allington Corp, 801 N. Pitt St, #707, Alexandria, VA 22314
DATE: 1977	AGE/GRADE LEVEL: 2-15 years
TEST INTENDS TO MEASURE:	Students' oral language ability; in particular, students are rated on a developmental hierarchy of language ability which runs from use of a single word or two through use of sentences and descriptions of various degrees of completeness, to telling a complete, well-organized, imaginative story, including implications, intentions, and predictions
TASKS INVOLVED:	Students tell a story about a picture; if response lags, prompting is done; three stories about three different pictures are collected
INDEX TO PART ONE:	Notions of Linguistic Complexity; Measures of Linguistic Creativity; Discrete Point vs. Integrative Testing Re Notions of Linguistic Complexity: The scoring scheme for this test is based on a nine-level scale of linguistic complexity which is said to represent the levels through which a child progresses to reach linguistic maturity. It is not clear what research this hierarchy is based on. One problem is that the lower levels appear to be measuring something very different from the higher ones. Up to about Level 5, students' responses are judged according to their formal characteristics: single word, complete sentence, verb of action, etc. In contrast, the highest levels (7, 8, and 9) depend on what we might call "literary" properties of the response: completeness, organization, imagination, and creativity. Two very different kinds of complexity are involved.

LOS NIETOS SCHOOL DISTRICT LANGUAGE DOMINANCE SURVEY

AUTHORS:	Los Nietos School District
AVAILABLE FROM:	José Garay, Title VII Director, Los Nietos School District, 8324 S. Westman Ave, Whittier, CA 90606
DATE: None	AGE/GRADE LEVEL: grades K-8
TEST INTENDS TO MEASURE:	Language used by various family members and in various social situations; ability to produce and understand English and Spanish
TASKS INVOLVED:	Parents answer questions about the language used by family members in different situations; students answer personal questions and perform activities directed by the examiner
INDEX TO PART ONE:	Self-Reported Data; Identical Items in Two Languages

MARYSVILLE TEST OF LANGUAGE DOMINANCE

AUTHORS:	Eleanor Thonis
AVAILABLE FROM:	Marysville Reading-Learning Center, Marysville Joint Unified School District, 1919 B St, Marysville, CA 95901
DATE: 1977	AGE/GRADE LEVEL: grades K-5
TEST INTENDS TO MEASURE:	Listening comprehension; speaking; reading (not reviewed); writing (not reviewed); cultural variables
TASKS INVOLVED:	Students perform activities directed by the examiner; students answer personal questions and perform a verbal task (counting); information is gathered from school records and from students about what language is used by various family members, the location of students' homes, and students' school history
INDEX TO PART ONE:	Self-Reported Data; Identical Items in Two Languages

MAT-SEA-CAL ORAL PROFICIENCY TEST

AUTHORS:	Joseph H. Matluck, Betty Mace-Matluck
AVAILABLE FROM:	Seattle Public Schools, 815 Fourth Ave North, Seattle, WA 98109
DATE: 1974	AGE/GRADE LEVEL: grades K-4
TEST INTENDS TO MEASURE:	Listening comprehension: grammar, vocabulary; speaking (repetition): pronunciation, grammar; speaking: pronunciation, grammar, vocabulary
TASKS INVOLVED:	Students identify pictures corresponding to sentences or groups of sentences spoken by the examiner; students repeat sentences spoken by the examiner; students answer questions posed by the examiner, who provides the first word or two of the desired response
INDEX TO PART ONE:	Tasks that Reduce to a Vocabulary Test; Mimicry; Passive Comprehension; The Value of Phonology or Pronunciation Tests; Ellipsis and Complete Sentences; Identical Items in Two Languages; Discrete Point vs. Integrative Testing; Correspondence between Test Content and Acquisition Order

Re Tasks that Reduce to a Vocabulary Test: In the "Listening Comprehension" section, students identify pictures corresponding to the examiner's sentences. The problem with such tests, as argued in this section, is that some items may not test sentence or structure comprehension, but simply vocabulary. This test does attempt to distinguish for each item, whether that item tests vocabulary or some aspect of structure, but the information is not used in scoring.

MICHIGAN ORAL LANGUAGE PRODUCTIVE TEST

AUTHORS:	John C. Larson
AVAILABLE FROM:	Modern Language Association, Publications Center, 62 Fifth Ave, New York, NY 10011
DATE: 1970	AGE/GRADE LEVEL: 4-6 years
TEST INTENDS TO MEASURE:	Students' ability to produce standard English grammar and pronunciation; the test is intended for speakers of ESL or for speakers of nonstandard English dialects
TASKS INVOLVED:	Students complete statements begun by the examiner, describing pictures; students answer questions about the pictures; students ask questions when instructed to by the examiner; in each case, the examiner says the first word or two of the desired response
INDEX TO PART ONE:	Ellipsis and Complete Sentences; Discrete Point vs. Integrative Testing; Correspondence between Test Content and Acquisition Order

MORENO ORAL ENGLISH PROFICIENCY PLACEMENT TEST

AUTHORS:	Steve Moreno
AVAILABLE FROM:	Moreno Educational Co, 7050 Belle Glade Ln, San Diego, CA 92119
DATE: 1974	AGE/GRADE LEVEL: 4-20 years
TEST INTENDS TO MEASURE:	Placement and achievement with respect to a particular English curriculum; also, overall English proficiency or degree of bilingualism
TASKS INVOLVED:	Students answer questions posed by the examiner; students repeat what the examiner says; students ask certain questions when the examiner instructs them to (most questions concern a set of pictures)
INDEX TO PART ONE:	Mimicry; Inferring Lack of Control from Lack of Performance; Understanding the Test Task; Ellipsis and Complete Sentences; Identical Items in Two Languages; Discrete Point vs. Integrative Testing; Correspondence between Test Content and Acquisition Order

Re Mimicry: A few items on this test are essentially mimicry items. For example, Question: *Tell me that he is not a girl*; suggested answer: *He is not a girl*. Question: *Tell me that you don't want any candy*; suggested answer: *I don't want any candy*. These items share a certain unnaturalness with the mimicry tasks imposed on other tests. Normal conversation does not include these kinds of questions.

Re Inferring Lack of Control from Lack of Performance: To get credit for a correct answer, students must produce the crucial

portion of the suggested answer (the portion underlined in the test booklet). However, for some items, a correct, appropriate, and grammatical answer might fail to include what the item intended to elicit. Under these circumstances, we cannot conclude that students do not control the expression being tested.

MORENO QUICK LANGUAGE ASSESSMENT INVENTORY

AUTHORS:	Steve Moreno
AVAILABLE FROM:	Moreno Educational Co, 7050 Belle Glade Ln, San Diego, CA 92119
DATE: 1974	AGE/GRADE LEVEL: grades K-6
TEST INTENDS TO MEASURE:	Students' language background for determining English/Spanish dominance
TASKS INVOLVED:	Parents answer questions about the students' native language and education, about the language used at home, and about parents' birthplace and education
INDEX TO PART ONE:	Self-Reported Data

NORTHWESTERN SYNTAX SCREENING TEST

AUTHORS:	Laura L. Lee
AVAILABLE FROM:	Department of Communicative Disorders, Northwestern University, Evanston, IL 60201
DATE: 1971	AGE/GRADE LEVEL: 3-8 years
TEST INTENDS TO MEASURE:	Understanding and production of English speech sounds; vocabulary; comprehension and production of English grammatical forms. [This test is intended as a screening instrument for language development in native English-speaking children. It is not intended to provide a linguistically detailed diagnosis or to assess language development in bilinguals or speakers of nonstandard dialects. However, "if an examiner wished to know how well the children of another dialect group used standard American dialect in reception and expression, it would be useful for that purpose (Manual)."]
TASKS INVOLVED:	Students identify pictures corresponding to sentences spoken by the examiner; the examiner presents two sentences orally, points to a picture, and students repeat the corresponding sentence
INDEX TO PART ONE:	Mimicry; Passive Comprehension; Discrete Point vs. Integrative Testing

ORAL LANGUAGE DOMINANCE MEASURE

AUTHORS:	Department of Curriculum and Staff Development, El Paso Public Schools
AVAILABLE FROM:	Primary Acquisition of Languages: A Dual Language Program, El Paso Public Schools, Department of Curriculum and Staff Development, El Paso, TX 79998

DATE: 1978	AGE/GRADE LEVEL: grades K-3
TEST INTENDS TO MEASURE:	Grammatical and appropriate language use; total number of words produced
TASKS INVOLVED:	Students answer questions posed by the examiner about a set of pictures
INDEX TO PART ONE:	Inferring Lack of Control from Lack of Performance; Counting Linguistic Elements; Ellipsis and Complete Sentences; Metalinguistic Instructions; Identical Items in Two Languages; Discrete Point vs. Integrative Testing; Correspondence between Test Content and Acquisition Order

ORAL LANGUAGE EVALUATION

AUTHORS:	Nicholas J. Silvaroli, Jann T. Skinner, J.O. Maynes, Jr.
AVAILABLE FROM:	EMC Corp, 180 East Sixth St, St. Paul, MN 55101
DATE: 1977	AGE/GRADE LEVEL: Not indicated
TEST INTENDS TO MEASURE:	Oral language ability (in particular, students are rated on a hierarchy of language development that runs from simple labeling, through more complex syntactic forms, to abstract story telling)
TASKS INVOLVED:	Students tell stories about two pictures; the examiner prompts if necessary
INDEX TO PART ONE:	Inferring Lack of Control from Lack of Performance; Understanding the Test Task; Measures of Linguistic Creativity; Discrete Point vs. Integrative Testing Re Inferring Lack of Control from Lack of Performance: On this test, students tell a story which is evaluated according to an assumed developmental hierarchy presented in the Teacher's Manual. The kind of performance rated highest includes specification of cause and effect, mood, emotion, etc. Students' failure to include these features in the story might result from factors other than not being able to specify such features.

OTT TEST OF ORAL LANGUAGE: ENGLISH AND SPANISH

AUTHORS:	Elizabeth Ott
AVAILABLE FROM:	Educational Testing Service, ETS Test Collection, Princeton, NJ 08540 (microfiche)
DATE: 1970	AGE/GRADE LEVEL: Not indicated
TEST INTENDS TO MEASURE:	The pronunciation of English sounds, especially those that may be difficult for native Spanish speakers; the fluency, intonation, and grammatical complexity of students' speech
TASKS INVOLVED:	Students report sentences spoken by the examiner; students describe pictures
INDEX TO PART ONE:	Notions of Linguistic Complexity; Mimicry; Inferring Lack of Control from Lack of Performance; Self-Reported Data; The Value of Phonology or Pronunciation Tests; Measures of Linguistic Creativity; Understanding the Test Task; Ellipsis and Complete Sentences; Identical Items in Two Languages

Re Measures of Linguistic Creativity: Some of the questions posed by the test require students to exercise some degree of imagination in their answers. For example, students are shown a picture of girls watching a game and must answer the question, *Why are the girls watching?* Some students may not be as imaginative in their answers or as interested in the test task as others, and may not produce elaborate or linguistically complex answers. According to the scoring strategy, these students would not do as well on these items.

PICTORIAL TEST OF BILINGUALISM AND LANGUAGE DOMINANCE

AUTHORS:	Darwin Nelson, Michael J. Fellner, C.L. Norrell
AVAILABLE FROM:	Texas Testing Services, Inc, 401 Poenisch, Corpus Christi, TX 78412
DATE: 1975	AGE/GRADE LEVEL: grades preK-2
TEST INTENDS TO MEASURE:	Vocabulary; oral language ability, judged for grammatical correctness and completeness
TASKS INVOLVED:	Students identify pictures in their choice of either English or Spanish, and then are asked to identify the pictures in the language they did not use the first time; students tell a story about a picture, first in one language, then the other; probing is done, if necessary
INDEX TO PART ONE:	The Value of Vocabulary Tests; Measures of Linguistic Creativity; Ellipsis and Complete Sentences; Metalinguistic Instructions; Identical Items in Two Languages; Discrete Point vs. Integrative Testing

ROBSTOWN INDEPENDENT SCHOOL DISTRICT ORAL LANGUAGE INVENTORY

AUTHORS:	Irma Garcia, Charlene Washburn
AVAILABLE FROM:	Bilingual Education Program, Robstown Independent School District, 101 West Ave E, Robstown, TX 78380
DATE: 1970	AGE/GRADE LEVEL: grades K-3
TEST INTENDS TO MEASURE:	Language used or preferred in particular social situations
TASKS INVOLVED:	Students answer questions about the language they prefer to use with particular people and in particular situations
INDEX TO PART ONE:	Tests that Require Literacy Skills; Self-Reported Data

SHORT TEST OF LINGUISTIC SKILLS

AUTHORS:	Department of Research and Evaluation, Chicago Board of Education
AVAILABLE FROM:	Department of Research and Evaluation, Chicago Board of Education, Rm 215, 2021 N. Burling St, Chicago, IL 60641
DATE: 1976	AGE/GRADE LEVEL: 8-13 years

TEST INTENDS TO MEASURE: Listening; reading (not reviewed); writing (not reviewed); speaking

TASKS INVOLVED: *Listening:* students choose words spoken by the examiner from a written multiple-choice list; write words, phrases, and sentences from dictation; answer yes/no questions; follow directions presented orally, many of which require writing. *Speaking:* students answer personal and general questions; describe objects and activities; answer questions about Chicago; describe a picture in five complete sentences

INDEX TO PART ONE: Tests that Require Literacy Skills; Notions of Linguistic Complexity; Measures of Linguistic Creativity; Ellipsis and Complete Sentences; Discrete Point vs. Integrative Testing

Re Measures of Linguistic Creativity: In scoring the speaking section, judgments must be made about "major" and "minor" grammatical errors. No useful assistance is provided for distinguishing between major and minor errors. Some attempt should be made to define what is considered a major versus a minor grammatical error, and indeed what is considered a grammatical error. Many stylistic and colloquial deviations from formally "correct" English should probably not be considered errors for purposes of an English proficiency test.

SHUTT PRIMARY LANGUAGE INDICATOR TEST

AUTHORS: D.L. Shutt

AVAILABLE FROM: Webster Division, McGraw-Hill, 8171 Redwood Hwy, Novato, CA 94947

DATE: 1976 AGE/GRADE LEVEL: grades K-6 (listening and verbal fluency)
grades 3-6 (reading comprehension and grammar)

TEST INTENDS TO MEASURE: Listening comprehension; verbal fluency; reading comprehension and grammar (not reviewed)

TASKS INVOLVED: Students identify pictures corresponding to what the examiner says; students describe pictures and answer specific questions about them

INDEX TO PART ONE: Tasks that Reduce to a Vocabulary Test; Passive Comprehension; Understanding the Test Task

Re Passive Comprehension: The "Listening Comprehension" section of this test seeks to test students' proficiency without actually making them produce speech.

SKOCZYLAS ENGLISH PHONEMIC UNIT PRODUCTION TEST

AUTHORS: Rudolph V. Skoczylas

AVAILABLE FROM: R.V. Skoczylas, 7649 Santa Inez Ct, Gilroy, CA 95020

DATE: 1972 AGE/GRADE LEVEL: 5-adult

TEST INTENDS TO MEASURE: Correct pronunciation of English speech sounds

TASKS INVOLVED: Repeating sentences spoken by the examiner

INDEX TO PART ONE: Mimicry; Tasks that Reduce to a Vocabulary Test; The Value of Phonology or Pronunciation Tests

SKOCZYLAS HOME BILINGUAL USAGE ESTIMATE

AUTHORS:	Rudolph V. Skoczylas
AVAILABLE FROM:	R.V. Skoczylas, 7649 Santa Inez Ct, Gilroy, CA 95020
DATE: 1971	AGE/GRADE LEVEL: Not indicated
TEST INTENDS TO MEASURE:	Students' linguistic background; language used by students' various family members
TASKS INVOLVED:	Adults answer questions about language use between students and family members and between other family members
INDEX TO PART ONE:	Self-Reported Data

SPANISH-ENGLISH DOMINANCE ASSESSMENT TEST

AUTHORS:	Bernard Spolsky, Penny Murphy
AVAILABLE FROM:	Educational Testing Service, ETS Test Collection, Princeton, NJ 08540 (microfiche)
DATE: 1972	AGE/GRADE LEVEL: grades 1-2
TEST INTENDS TO MEASURE:	Oral language production; vocabulary; use and understanding of English and Spanish in particular social situations
TASKS INVOLVED:	Students answer personal questions, questions about their own language use and understanding; name objects that might be found in particular places; produce connected speech describing a picture
INDEX TO PART ONE:	Self-Reported Data; The Value of Vocabulary Tests; Metalinguistic Instructions; Identical Items in Two Languages

SPANISH/ENGLISH LANGUAGE PERFORMANCE SCREENING

AUTHORS:	Southwest Educational Development Laboratory
AVAILABLE FROM:	CTB/McGraw-Hill, Del Monte Research Park, Monterey, CA 93940
DATE: 1976	AGE/GRADE LEVEL: 4-5 years
TEST INTENDS TO MEASURE:	Oral language ability, judged according to number, length, and complexity of responses
TASKS INVOLVED:	Students answer personal questions; name objects; follow oral directions; describe objects and pictures
INDEX TO PART ONE:	Notions of Linguistic Complexity; Metalinguistic Instructions; Identical Items in Two Languages; Discrete Point vs. Integrative Testing

Re Notions of Linguistic Complexity: As part of the procedure for determining language dominance, the number, length, and complexity of responses on the English section are compared with the responses on the Spanish section. No guidelines are provided for making this comparison. Judging the complexity of a response is likely to be a difficult and subjective matter, and comparing the complexity of responses in two different languages even more difficult.

Re Identical Items in Two Languages: Questions asked on each

language section are parallel, but not mere translations. This test avoids the unnaturalness of asking students the same question twice.

TEST OF GRAMMATICALLY CORRECT SPANISH/ENGLISH

AUTHORS:	Las Cruces Public Schools
AVAILABLE FROM:	Margarita López de Mestas, Coordinator, Bilingual Education Project, Las Cruces Public Schools, 301 W. Amador Ave, Las Cruces, NM 88001
DATE: None	AGE/GRADE LEVEL: grades K-4
TEST INTENDS TO MEASURE:	Correctness of grammatical form and spelling in a written story (not reviewed); average oral sentence length, number of uncommon words, total number of syllables used: grammatical complexity and correctness; pronunciation
TASKS INVOLVED:	Students answer two questions in Spanish, then two different questions in English; probing is done if necessary (oral test)
INDEX TO PART ONE:	Notions of Linguistic Complexity; Inferring Lack of Control from Lack of Performance; The Value of Phonology or Pronunciation Tests; Counting Linguistic Elements; Identical Items in Two Languages; Discrete Point vs. Integrative Testing

INDEX

BAHIA ORAL LANGUAGE TEST 29, 31, 36
BASIC INVENTORY OF NATURAL LANGUAGE 18, 19, 22, 25, 27, 28, 36
BILINGUAL SYNTAX MEASURE 23, 25, 29, 31, 37
BROOKS COUNTY LANGUAGE USAGE INVENTORY 12, 13, 37

CASO TEST FOR NON-ENGLISH SPEAKING STUDENTS 17, 22, 29, 30, 37
CHICAGO *See* SHORT TEST OF LINGUISTIC SKILLS
COMPREHENSION OF ORAL LANGUAGE 28, 29, 38
CRANE ORAL DOMINANCE TEST 8, 20, 24, 38

DEL RIO LANGUAGE SCREENING TEST 9, 10, 29, 39

EL PASO *See* ORAL LANGUAGE DOMINANCE MEASURE
ENGLISH AS A SECOND LANGUAGE ACHIEVEMENT TESTS--LEVEL I 10, 22, 29, 30, 39

FLEXIBILITY TEST TO MEASURE DOMINANCE IN SPANISH-ENGLISH BILINGUALS 25, 40

HARRIS ARTICULATION TEST 10, 24, 41
HAYWARD LANGUAGE DOMINANCE INDICATOR 12, 13, 14, 19, 41

ILYIN ORAL INTERVIEW 17, 20, 21, 23, 29, 30, 41

JAMES LANGUAGE DOMINANCE TEST 14, 24, 25, 42

LANGUAGE ASSESSMENT BATTERY 8, 14, 15, 23, 29, 30, 42
LANGUAGE ASSESSMENT SCALES 10, 11, 14, 28, 29, 43
LANGUAGE DOMINANCE CRITERIA 12, 28, 43
LANGUAGE DOMINANCE INDICATOR *See* HAYWARD
LANGUAGE DOMINANCE SURVEY 12, 27, 44
LANGUAGE FACILITY TEST 17, 18, 27, 28, 44
LOS NIETOS SCHOOL DISTRICT LANGUAGE DOMINANCE SURVEY 12, 45

MARYSVILLE TEST OF LANGUAGE DOMINANCE 12, 45
MAT-SEA-CAL ORAL PROFICIENCY TEST 10, 23, 29, 30, 45
MICHIGAN ORAL LANGUAGE PRODUCTIVE TEST 22, 23, 29, 31, 46
MILWAUKEE *See* ENGLISH AS A SECOND LANGUAGE ACHIEVEMENT TESTS--LEVEL I
MORENO ORAL ENGLISH PROFICIENCY PLACEMENT TEST 20, 21, 29, 30, 46
MORENO QUICK LANGUAGE ASSESSMENT INVENTORY 12, 47

NORTHWESTERN SYNTAX SCREENING TEST 10, 29, 47

ORAL LANGUAGE DOMINANCE MEASURE 11, 19, 23, 24, 25, 26, 29, 31, 47
ORAL LANGUAGE EVALUATION 17, 27, 28, 48
OTT TEST OF ORAL LANGUAGE: ENGLISH AND SPANISH 9, 10, 11, 15, 22, 48

PICTORIAL TEST OF BILINGUALISM AND LANGUAGE DOMINANCE 13, 14, 17, 22, 24, 25, 27, 28, 49

ROBSTOWN INDEPENDENT SCHOOL DISTRICT ORAL LANGUAGE INVENTORY 12, 49

SHORT TEST OF LINGUISTIC SKILLS 7, 22, 28, 49
SHUTT PRIMARY LANGUAGE INDICATOR TEST 8, 21, 50
SKOCZYLAS ENGLISH PHONEMIC UNIT PRODUCTION TEST 10, 15, 50
SKOCZYLAS HOME BILINGUAL USAGE ESTIMATE 12, 51
SPANISH-ENGLISH DOMINANCE ASSESSMENT TEST 12, 14, 26, 51
SPANISH/ENGLISH LANGUAGE PERFORMANCE SCREENING 24, 28, 51

TEST OF GRAMMATICALLY CORRECT SPANISH/ENGLISH 11, 15, 19, 28, 52

Cecilia Freeman is a Research Associate at the Center for Applied Linguistics, where she has worked on issues involving language in education. She is author of *Readability and Text Structure: A View from Linguistics* (Center for Applied Linguistics, 1978) and, with Thomas G. Dieterich, of the *Arlington County Oral Production Test* (Arlington, Virginia Public Schools, 1978) and other works in education and linguistics.

Thomas G. Dieterich (Ph.D., Yale University) is Assistant Professor of English at Portland State University, Portland, Oregon. He is co-author, with Cecilia Freeman and Peg Griffin, of *Assessing Comprehension in a School Setting* (Center for Applied Linguistics, 1978) and of other works on linguistics and on language and education.

LANGUAGE IN EDUCATION: THEORY AND PRACTICE

Volume 1 (1977-78)

1. Directory of Foreign Language Service Organizations, by Sophia Behrens. $3.95. ED 153 503
2. The Linguist in Speech Pathology, by Walt Wolfram. $2.95. ED 153 504
3. Graduate Theses and Dissertations in English as a Second Language: 1976-77, by Stephen Cooper. $2.95. ED 153 505
4. Code Switching and the Classroom Teacher, by Guadalupe Valdés-Fallis. $2.95. ED 153 506
5. Current Approaches to the Teaching of Grammar in ESL, by David M. Davidson. $2.95. ED 154 620
6. From the Community to the Classroom: Gathering Second-Language Speech Samples, by Barbara F. Freed. $2.95. ED 157 404
7. Kinesics and Cross-Cultural Understanding, by Genelle G. Morain. $2.95. ED 157 405
8. New Perspectives on Teaching Vocabulary, by Howard H. Keller. $2.95. ED 157 406
9. Teacher Talk: Language in the Classroom, by Shirley B. Heath. $2.95. ED 158 575
10. Language and Linguistics: Bases for a Curriculum, by Julia S. Falk. $2.95. ED 158 576
11. Teaching Culture: Strategies and Techniques, by Robert C. Lafayette. $2.95. ED 157 407
12. Personality and Second Language Learning, by Virginia D. Hodge. $2.95. ED 157 408

Volume 2 (1978-79)

13. Games and Simulations in the Foreign Language Classroom, by Alice C. Omaggio. $[illegible].95.
14. Problems and Teaching Strategies in ESL Composition, by Ann Raimes. $2.95.
15. Graduate Theses and Dissertations in English as a Second Language: 1977-78, by Stephen Cooper. $2.95.
16. Foreign Languages, English as a Second/Foreign Language, and the U.S. Multinational Corporation, by Marianne Inman. $4.95.
17. Testing Oral Communication in the Foreign Language Classroom, by Walter H. Bartz. $2.95.
18. Intensive Foreign Language Courses, by David P. Benseler and Renate A. Schulz. $5.95.
19. Evaluating a Second Language Program, by Gilbert A. Jarvis and Shirley J. Adams. $2.95.
20. Reading a Second Language, by G. Truett Cates and Janet K. Swaffar. $2.95.
21. Chicano English, by Allan A. Metcalf. $2.95.
22. Adult Vocational ESL, by Jo Ann Crandall. $5.95.

Volume 3 (1979-80)

23. A Linguistic Guide to English Proficiency Testing in Schools, by Cecilia Freeman and Thomas G. Dieterich. $5.95.

To subscribe to the complete series of publications, write to the Publications Department, Center for Applied Linguistics, 1611 North Kent Street, Arlington, Virginia 22209. The subscription rate is $32.00 per volume. Titles may also be ordered individually; add $1.75 postage and handling. Virginia residents add 4% sales tax. All orders must be prepaid.